RESEARCH TRIANGLE INSTITUTE

March 1981
Draft of Final

1864/14/03 - 01I

National Soil Monitoring Program

by

Roy Whitmore
Martin Rosenzweig
John Hines

Research Triangle Institute
Research Triangle Park,
North Carolina 27709

Contract No. 68-01-5848

Task Manager: William Smith
Project Officer: Ann Carey

Design and Development Branch
Exposure Evaluation Division
Office of Pesticides and Toxic Substances
Environmental Protection Agency
Washington, D.C. 20460

Disclaimer

This document is a preliminary draft. It has not been released formally by the Office of Testing and Evaluation, Office of Pesticides and Toxic Substances, U.S. Environmental Protection Agency, and should not at this stage be construed to represent Agency policy. It is being circulated for comments on its technical merit and policy implications.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

1. Introduction

The purpose of the review of the National Soils Monitoring Program (NSMP) is to:

a) Describe the network,
b) Assess it current effectiveness,
c) Provide design options.

The NSMP has two components, the Urban Soils Network (USN) and the Rural Soils Network (RSN). Its purpose has been to monitor pesticide residues in soils in the conterminous United States.

The USN will be reviewed in a later report.

This report considers the RSN review which represents a major and time-consuming effort. It embraces the assembly and review of design documents, the correspondence files and memoranda relating to operational activities, and the computer data files including editing and correcting data entries where necessary. It also includes analyses of the data using the sampling weights developed during the establishment of the structure of the survey design.

This report contains a brief and complete description of the statistical design of the RSN, and its parent the CNI. It is therefore a valuable asset in understanding, analyzing or modifying the soil monitoring efforts of the federal government.

1.1 General Description of the Program

The National Soil Monitoring Program consists of two networks: (1) the Urban Soils Network and (2) the Rural Soils Network. The Rural Soils Network is a probability subsample of the 1967 Conservation Needs Inventory sample. The area sampled by the Rural Soils Network includes all of the conterminous United States except for areas considered to be urban in character. These urban areas are monitored by the Urban Soils Network, which consists of a stratified sample.

nice - 1.2.1 General Considerations

The fact the Rural Soils Network (RSN) is a probability sample makes possible valid statistical inferences to the population sampled, namely all rural soils of the conterminous United States. Moreover, inferences are possible for all reasonably large geographic areas within the United States, for example cropping regions and larger States. Some State exclusions must be noted in analyzing the data.

The operational design of the RSN makes possible some interesting statistical analyses. Because soil and crop specimens are obtained simultaneously at harvest time from matched sites, the relationship between pesticide levels in soils and harvested crops can be analyzed.

Also, since some sites were sampled at a four year interval, trends in pesticide residue levels can be investigated.

1.2.2 The Sampling Design

The Rural Soil Network (RSN) is a probability sample of 10-acre sites from the population of all rural land areas in the conterminous United States. Each 10-acre site is located by a probability subsample of the data points of the 1967 Conservation Needs Inventory (CNI). The CNI, in turn, is a probability sample of all rural land areas in the conterminous United States.

The CNI is a stratified random sample of primary sampling units (PSU's) from each county of the conterminous United States, except for those counties strictly metropolitan in character. The standard size of the PSU's was 160 acres, although 40-acre, 100-acre, and 640-acre PSU's were not uncommon. The standard sampling rate was two percent, however this rate was increased or decreased in order to either provide estimates of nearly equal precision for all counties and to oversample areas of special interest. The sampling rates varied within strata from less than one percent to approximately thirty-two percent.

In the CNI, data were collected for each of a series of points at every CNI sample site. The land use data collected for each CNI sampling point was used to classify the point as either a cropland point or a noncropland point. The sampling design of the RSN specified that 0.025 percent of the cropland and 0.0025 percent of the noncropland of the rural conterminous United States would be sampled. A subsample of the CNI cropland sampling points was selected and used to locate the RSN cropland sample sites. The RSN noncropland sample sites were located by a subsample of the CNI noncropland sampling points.

The operational design of the Rural Soils Network (RSN) specifies that each cropland site be randomly designated as a first-year, second-year, third-year, or fourth-year cropland site, such that one-fourth of the cropland sites in each State will be sampled each fiscal year. Noncropland sites were handled in the same manner. Specimens were to be collected at each site no less than once every four years and not more than once per year. Soil specimens were obtained by compositing fifty soil cores, 2-inches in diameter by 3-inches in depth. Cropland specimens were to be obtained immediately before or at harvest time.

1.3 Alternate Survey Designs

1.3.1 Design Option One

A minimal change alternative would be to subsample the current RSN. This option mainly addresses the problem of the cost of the RSN although the need for national and regional estimates is also considered. [Any need to eliminate reliance upon the 1967 CNI is not addressed.]

This option does offer the advantage that it can be quickly and easily implemented, possibly while other alternatives are under development.

Replicate subsamples are recommended if this option is to be implemented, even if it is only on a temporary basis. For example, if 50 percent of the RSN sites are to be surveyed, five subsamples that each comprise a 10 percent subsample can be used. At least five replicate subsamples should be selected. The use of replicate subsamples makes it possible to estimate sample variances easily by using the theory of replicate subsamples.

It may also be useful to select the subsamples at different rates within domains of interest. Identification of strata of special interest within the domains just considered can be used to increase the possibility of finding toxic substance residues.

1.3.2 Design Option Two

A design analogous to the design that produced the present RSN sample can be based upon the 1982 National Resources Inventory (NRI). Use of the 1982 NRI will provide up-to-date land use information. A subsampling procedure to obtain adequate precision at minimum cost is proposed. This can be accomplished by identifying areas where toxic residues are likely to be found and giving these areas a greater probability of being selected for the RSN sample.

It is suggested that counties be used as primary sampling units for the second phase sample. The data from the present RSN indicates that counties are generally heterogeneous with respect to toxic residues. Thus, it would be advantageous to select relatively few counties with a larger number of sample sites. The use of counties as PSU's will reduce travel costs associated with data collection. More importantly, smaller areas like counties can be effectively stratified into areas where toxic residues are likely to be found.

The RSN sample sites are to be located at NRI sample points. Sample counties are selected from the counties in the NRI sample, so that counties where toxic substance residues are likely have a greater chance of selection. Thus, it is suggested that counties be selected with probability proportional to size (PPS), where the size measure is a measure of the likelihood of finding toxic residues.

Efficient sampling within the selected counties can result from careful stratification. The NRI sampling points within a county are first stratified into cropland points and noncropland points, to insure adequate representation of each of these land types and because agricultural chemical residues are more likely to be found in cropland. Local land use characteristics can be used to further stratify both the cropland points and the noncropland points.

1.3.3 Design Option 3

Review of the data indicates large numbers of zero valued observations, and relatively few positive observations. This analytic challenge has been discussed elsewhere [See Lucas et al, Recommendations for the National Surface Water Monitoring Program for Pesticides. Report No.

RTI/1864/01-02I]. The conclusion of that analysis was that the appropriate measures of "level" are:

(1) The proportion of positive detections, i.e., the relative frequency of last stage sampling units positive for the substance(s) under investigation, and

(2) The proportion of sampling units containing concentrations of substance above some specified level. This level may signal the existence of an undesirable situation.

The proposed design is a two-stage area probability sample with stratification of the sampling units at each level. The first stage or primary sampling units (PSU's) are counties. Geographic stratification is provided by the four Census Regions. Allocation of PSU's to these regions is in proportion to the land area eligible for the study. Using additional variables to allocate the sample is unlikely to be useful at this level due to the variety of land use within each Census Region. The eligible land area is currently defined by the membership requirements of the RSN and USN. It may be advantageous from administrative as well as fiscal and statistical grounds to combine the activities of the soil networks, and consider SMSA counties as a stratum within the survey. This point requires further review, however initial investigation suggests savings are likely.

With the extension of monitoring responsibility from pesticides to toxic substances in general, some revision of the approach is indicated. The following stratification variables are therefore proposed for the PSU's in addition to the geographic stratification above:

(1) Land area,
(2) Population density,
(3) Agricultural activity, and
(4) Industrial activity.

Second stage sampling units (SSU's) are 10-acre plots. These are proposed as the final stage units or analysis units on the assumption that they are sufficiently homogeneous that the effects of subsampling are negligible . This is a verifiable proposition. The problem with SSU's this small is the ability to locate them in the field. The lack of identifiable boundaries renders exactly locating them most difficult. To ease this difficulty, enumeration districts (ED's) are proposed as readily identified segments. The problem is reduced to locating the SSU within the ED, or any suitable subsegment chosen to facilitate the task.

SSU's will be allocated equally to PSU's. A detailed field-use protocol will locate the specimen for collection, leaving the minimum of discretion for the field personnel in the selection of these sites. The protocol will specify a grid locating multiple specimen collection sites. The soil collected in a given plot would be composited, unless the homogeneity of the 10-acre plot is under investigation.

1.4 Present Network Operations

1.5 Alternate Operational Design

The operational design of the Rural Soils Network (RSN) was well conceived for monitoring agricultural pesticides and herbicides. However, much pesticide and herbicide residue may often be leached out of, or vaporized from, the cropland soil by harvest time.

1.6 Recommended Modifications

1.7 Statistical Findings

Several types of analyses are of interest for the RSN data, notably:

(1) Estimation of base levels for residues of toxic substances,
(2) Estimation of changes in mean levels of toxic substance residues from the first round to the second round of data collection, and
(3) Estimation of relationships between soil and crop residue levels.

The reason for analyzing the RSN data in this study was to obtain a measure of the degree of precision that could be obtained for analysis of residue data based upon the present data. It was decided that estimation of base levels of residues would be sufficient. In particular, estimation of levels was undertaken for the first round soil data only.

It was found that the data values for most compounds were predominantly zero. The predominance of zero values in the residue data results in J-shaped distributions for the amount of residue detected for most compounds. This type of data presents some rather unique analysis problems. For example, the weighted mean of the raw data values has little meaning if most values are zero and a few are very large. Thus, some type of data transformation is generally required in order to obtain a meaningful analysis [See Lucas, et al, Recommendations for the National Surface Water Monitoring Program for Pesticides. Report No. RTI/1864/01-02I]. Ideally, each compound should be considered individually to determine an appropriate transformation, if any. Ubiquitous compounds like arsenic may not require transformation.

For analyses on the proportion scale, all data values above the minimum detectable level (MDL) were replaced by the value one. The weighted mean on this scale is a weighted estimate of the proportion of the sampled land area with a residue level in excess of the MDL. Since this scale was felt to be generally the most appropriate for analysis of the residue data, the standard error and the design effect for the estimated proportion were also computed.

Estimation of standard errors and design effects required that some strata be combined. Since it was not possible to account for all dimensions of the CNI stratification, the standard errors computed are undoubtedly conservative estimates. This results in similarly conservative interval estimates of the proportion of sampled areas where levels of the compound exceed the minimum detectable level (MDL).

The design effect is the ratio of the sample standard error to an estimate of what the standard error would have been if a simple random sample of the same size had been used, i.e.,

$$\text{DEFF} = \frac{\text{Estimated S.E. (For the design used)}}{\text{Estimated S.E. (Simple Random Sample)}}$$

Alternatively, the design effect can be thought of as the ratio of the actual sample size to the sample size that would be required to obtain an estimate with the same standard error based upon a simple random sample. Generally stratification decreases the design effect, while clustering increases it. Thus, since the CNI stratification can be used and there is no clustering of sample sites in the RSN sample, design effects less one would be expected. This would indicate that the design produced smaller standard errors than would a simple random sample of the same size. Many of the design effects shown in Tables 1.7.7 through 1.7.9 are indeed less than one. However, some design effects are substantially greater than one. It is not clear therefore that the CNI stratification was particularly advantageous for estimation of proportions of detections for toxic substance residues.

1.8 Capabilities for Special Studies

1.9 Toxic Substances Other than Pesticides in Soils

1.10 Implementation Plan for a New Survey Design of the Rural Soils Network

2.0 Evaluation of Chemical Analysis

Information on the quality of the pesticide data compiled by the NSMP is not currently available to users of the program's computer data file. Some measure of this quality is necessary for meaningful statistical evaluation of the data and practical interpretation of the results. To this purpose, a limited review of the current analytical methodology was conducted and information compiled on the accuracy (recoveries), precision (coefficient of variation) and minimum detectable levels of each of the pesticides monitored under the program where such information was available.

Over thirty toxic substances have been monitored under the NSMP including several chemical classes: 1) organochlorine pesticides; 2) PCBs*; 3) trifluraline; 4) organophosphorous pesticides; and 5) heavy metals. All analyses (~ 450 soil specimens/year) are carried out at the Toxicant Analysis Center, Bay St. Louis, Mississippi. However, heavy metals have not been analyzed in soil since 1979.

Nearly all procedures applied to the analysis of pesticides and PCBs in soil specimens used an initial extraction followed by column chromatography clean-up. Final quantitation of pesticides was carried out using external standard techniques with gas chromatography (GC). In general, confirmation of detected pesticides was performed by changing the selectivity of the GC column or detector. Each set of specimens was

*Polychlorinated Biphenyls

accompanied by a blank and one or more controls (fortified blanks) to check contamination and pesticide recoveries during the extraction, clean-up and GC analysis procedures.

Levels of heavy metals in soil specimens were determined using atomic absorption spectroscopy (AA). Plane AA was used for lead, cadmium and arsenic and the cold vapor techniques for mercury. No information was available on the current accuracy, precision and limits of detection.

Relatively little information was readily available on the current accuracy, precision and MDLs** for pesticides and PCBs in soil. Of particular interest are individual values for accuracy and precision for each pesticide in each of the specimen matrices (crops, water and sediment). An average of each of these values derived from replicate analysis over a period of time would also provide an indication of the method stability for a particular pesticide in a specific matrix. Recovery data for each pesticide was judged a reasonable indication of method accuracy since analytical results not corrected for recoveries and losses during the analysis can represent a significant contribution to error in the reported result where such recoveries are low.

Relatively little recovery and precision data were available at levels near the pesticide MDLs. It is particularly important that such data be provided to users of the computer data files since it represents the "worst" case in terms of the data quality.

Limited review of analytical methodology used in the NSMP and an attempt to compile data for the average accuracy, precision and MDL in soil for each toxic substance monitored under this program provide a basis for the following recommendations:

1. Accuracy (that is, recoveries) and precision data must be generated for all pesticides monitored in the NSMP. The data should be generated at two different levels (e.g., at the MDL and at ten times the MDL). The results for controls analyzed with each set of specimens would be the best means of providing this information since it is necessary that control data be made accessible to computer data file users in any event. Controls must be run with each set of specimens and should consist of a blank (unfortified soil free from the analytes of interest) and two fortified blanks (one fortified at the MDL and another at ten times the MDL). The analytical results for the controls should be reported on a separate form (especially designed for control data) and encoded such that there is a one-to-one association with the particular set of specimens with which they were analyzed. The encoding should allow later computer retrieval of control data for any particular specimen set or group of sets (for example, geographic area, over a specified period of time, or for a particular pesticide). The availability of this information in a retrievable form to data file users would provide the means for assessing data reliability now lacking. Further, any duplicate specimen analyses must be reported in the computer data file as they provide the best means of assessing

---

** Minimum detection levels

method precision on a continuous basis. Duplicate results must be specifically encoded such that they are retrieved as a group (e.g., all duplicates for a particular matrix and pesticide over a specified period of time) as well as with the initial analytical results for the specimen. The need to make routine control data available to program data file users cannot be overemphasized. This does not preclude the use of specialized controls (e.g., SPRMS); however, these results should also be included in the computer file encoded to allow facile retrieval both as a group and with their particular specimen set.

2. The pesticides included on the routine monitoring list must be reviewed on a regular basis and appropriate deletions or additions made. Specifically, the need for routine analysis of organophosphorous pesticides in soil should be reviewed as this class of compounds is known to be unstable and has seldom been reported in either soil or sediment. Once the baseline has been established for such compounds, three choices are possible: 1) cease to analyze for the compound(s) except under special circumstances (e.g., after a chemical spill or when contamination is suspected from a recent application); 2) analyze for the compound(s) on a more frequent basis; and 3) concentrate efforts on the analysis of degradation products of known toxicity where these exist. Decisions concerning the analysis of toxic substances under the NSMP should be based on information generated in other agency data files (e.g., USDA, USGS, etc.) as well as data generated within EPA.

3. Soil specimens should be characterized as to the percent carbon or percent inogranic residue. This information must be included on the report form (along with moisture content) as part of the specimen characterization (source). Significant trends may otherwise be missed with respect to the soil type and its effect on toxic substance accumulation, degradation and transport.

4. Control specimens (in the matrix of interest) should be included with any specimens either stored for extended periods or shipped to another site for analysis. This is particularly important for toxic compounds which are known to be unstable; i.e., organophosphorous pesticides. The results of these "storage controls" must also be included in the computer data file with appropriate encoding for specific retrieval.

5. Analytical methodology should be updated to include state-of-the-art capillary GC techniques. This would provide a higher degree of confidence in the resulting data through increased resolution and sensitivity. The use of higher resolution analytical techniques is a move toward the quantitation of PCBs (and technical chlordane) as their individual isomers. This approach is far more useful than the present method of attempting to identify patterns and averaging components, since the toxicity and biodegradation of the individual isomers are not identical.

6. The pesticide recoveries should be monitored for each specimen analyzed by initial fortification of the specimen with appropriate

compound(s). Subsequent analysis of the compound level should enable comparison of data between specimens with increased confidence that anomalous results will be detected. The use of internal standard quantitation techniques would normalize recoveries between specimens and should be considered.

7. Detailed information on all analytical procedures under the NSMP should be documented in one source. The procedures must then be maintained current with ongoing improvements and modifications made by the analytical laboratories. Such updating requires both flexibility and regular review by program management.

1. NATIONAL SOILS MONITORING PROGRAM

1.1 General Description of the Program

The National Soils Monitoring Program consists of two networks: 1) Urban Soils Network and 2) Rural Soils Network. The Rural Soils Network (RSN) is a two phase probability sample. The first phase sample was the 1967 Conservation Needs Inventory (CNI) sample. The RSN sample is a probability subsample from the ultimate sampling units of the 1967 CNI. The area sampled by the RSN includes all of the conterminous United States except for areas considered to be urban in character. These urban areas are monitored by the Urban Soils Network, which consists of a sample of the urban areas.

1.2 The Rural Soils Network (RSN) Survey Design

1.2.1 General Considerations

The fact that the Rural Soils Network (RSN) is a probability sample makes possible valid statistical inferences to the population sampled, namely all rural soils of the conterminous United States. Moreover, inferences are available for all reasonably large geographic areas within the United States, e.g., cropping regions and the larger States. However, the decision not to collect data in some States restricts the population for which inferences are valid.

The operational design of the RSN makes possible some interesting statistical analyses. Since soil and crop samples are obtained simultaneously at harvest time, the relationship between pesticide levels in soils and harvested crops can be analyzed. Also, since each sample site is sampled at four-year intervals, trends in pesticide residue levels can be investigated.

1.2.2 The Probability Sample Design

The Rural Soils Network (RSN) is a probability sample of 10-acre sites from the population of all rural land areas in the conterminous United States. Each 10-acre site is located by a point determined by a probability subsample of the data points of the 1967 Conservation Needs Inventory (CNI) which is, in itself, a probability sample of all rural land areas in the conterminous United States. Among the lands included in the CNI are the following: (a) privately owned land, both personal and corporate; (b) land owned by State and local governments; (c) land owned by the federal government; and (d) Indian land. Among the areas excluded are: Ponds and lakes of more than two acres, all streams, and urban or built-up areas.

1.2.2.1 The CNI survey

The 1967 CNI did not however map, that is, collect data for, federal noncroplands. This portion of the CNI was indefinitely postponed, although all federally owned rural land areas did receive their share of CNI primary sampling units. Federally owned cropland operated under lease or permit was, however, mapped by the 1967 CNI.

Urban or built-up areas excluded from the CNI have a specific definition and not all areas inside city and village limits are considered urban or built-up, whereas some areas outside city and village limits are. In particular, urban or built-up areas are defined as areas of 10 acres or more, consisting of residential sites, industrial sites (except strip mines, borrow and gravel pits), railroads, roadways, cemeteries, airports, golf courses, shooting ranges, institutional and public administration sites, and "similar kinds of areas."[1] The exclusion of urban or built-up areas (of 10 acres or more) from the CNI resulted in excluding of some counties that were strictly metropolitan in character.

The CNI sample sites were selected by the Statistical Laboratories at Cornell University and Iowa State University. The sampling sites for thirteen States in the northeastern United States were selected at Cornell. All other sampling sites were selected at Iowa State. A deeply stratified sampling design was used for the CNI. Counties were treated as strata within all States. Little more is known about the procedure used at Cornell, except that the standard sampling rate was about 2 percent and the standard size of a primary sampling unit (PSU) was 100 acres. The stratification used at Iowa State sometimes involved large scale geographic stratification between the State and county levels, e.g., a sandhills stratum was designated in Nebraska, and in many States irrigated areas were treated as a stratum.

The sampling procedure followed at Iowa State can best be understood by first considering the procedure most commonly employed in the States of the western United States that are divided into townships. A township is a 6 mile by 6 mile square of land (see figure 1.1). Each regular township contains 36 sections. This township consists of 6 rows, each containing 6 sections. Three geographical strata were formed from this township: 1) the first stratum was the northern 2 rows; 2) the second stratum was the middle 2 rows; and 3) the third stratum consisted of the 2 southernmost rows. Each stratum then contained 48 quarter-sections (160-acre square PSU's), from which a predetermined number of PSU's were randomly selected. The standard sampling rate for the 1967 CNI was the selection of one PSU from each stratum of 48 PSU's. Thus, the standard sampling rate was approximately 2 percent (1/48).

Estimates of nearly equal precision were desired for all counties. The sampling procedure just described was believed to provide sufficient precision for a county with 384 to 767 acres of inventory acreage. Thus, a sampling rate of less than 2% was used in some of the larger counties, and more than 2% in some of the smaller counties. The sampling rate was also generally increased in irrigated strata and other areas of special interest.

In order to increase the sampling rate from 2% to 4%, two quarter-sections were selected from each stratum, rather than one. However, a

---

[1] Basic Statistics -- National Inventory of Soil and Water Conservation Needs, 1967.

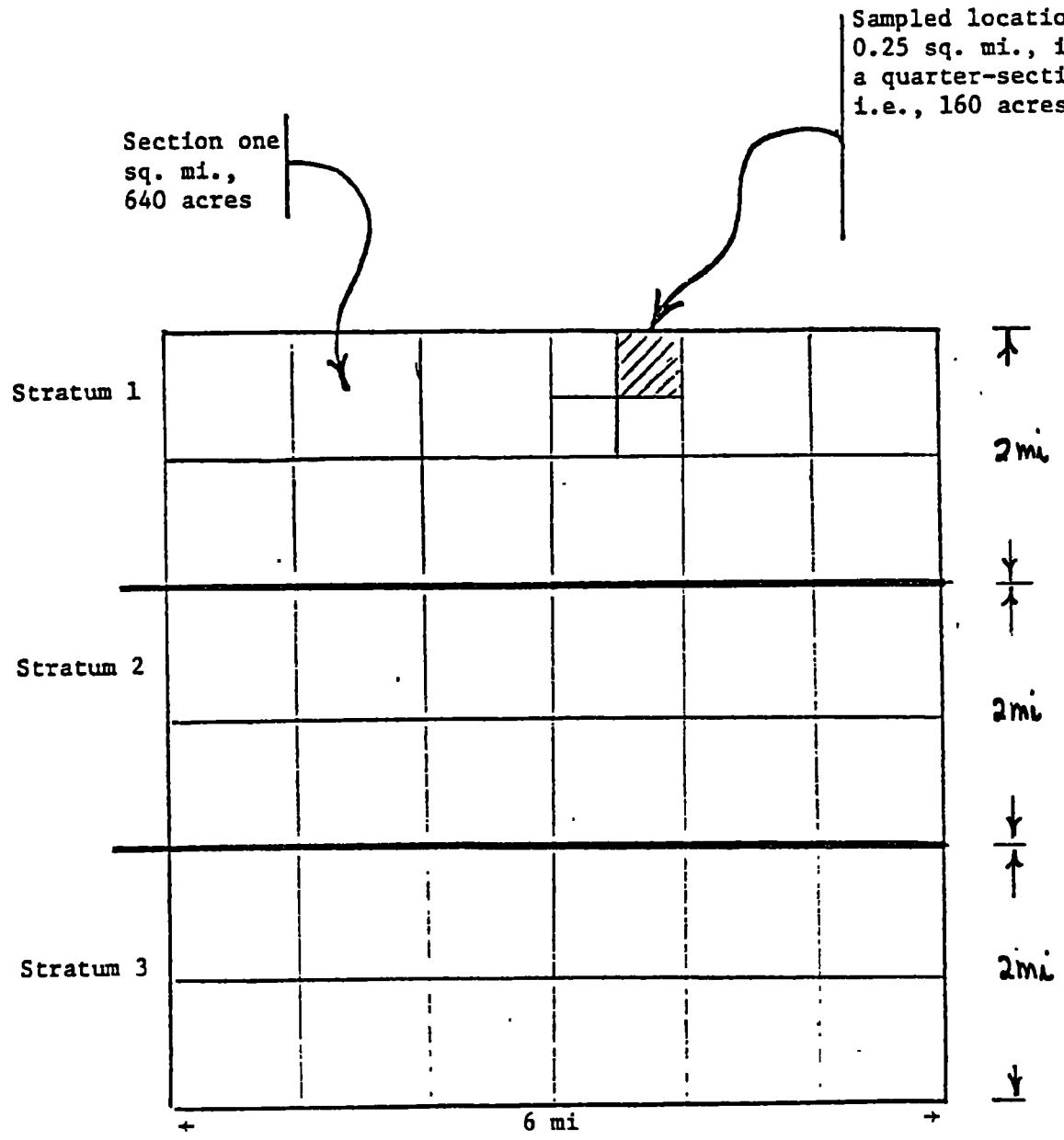


Figure 1.1 Typical Stratification of a Township

(Source: personal communication from Iowa State University, Statistical Laboratory).

decrease in the sampling rate from 2% to 1% was accomplished by changing the stratum size from 12 sections to 24 sections with one quarter-section being selected from each of the 24 section strata. Thus, a decrease in the sampling rate from 2% was accompanied by an increase in the stratum size.

It was also desirable at times to change the size of the CNI sampling site from the usual 160 acres. In some large counties in the western United States with large tracts of relatively homogeneous soil type and usage, CNI sample sites consisted of one section or 640 acres. In some highly developed agricultural areas of special interest, sites consisting of 40 acres, a sixteenth-section, were sometimes used because of considerable heterogeneity between fields.

The above considerations led to the establishment of Table 1.1 for the determination of a standard sampling rate based upon the inventory acreage of a county and the size of sampling unit to be used. The standard sampling rates shown in Table 1.1 were determined so that the relative precision of county level estimates would be constant, i.e. not dependent upon either county or sampling unit size. This table was not strictly adhered to, however.

The sampling procedure just described was used in all States samples designed at Iowa State. Township and section boundaries were artificially imposed upon counties that were not already surveyed into such divisions. Whenever possible, township and section boundaries were made to follow lines of longitude and latitude in the same manner as in sectionized States.

Many counties are not regular in shape so that there were often partial townships, strata, and sections around their borders. Sections around such borders were included in the sampling frame only if at least part of the section was in the county being sampled. Such sections were then grouped into strata for sample selection. The strata were usually composed of twelve sections each, just as twelve sections form one stratum in the standard sampling scheme depicted in Figure 1.1. Any sampling units that fell outside the county of interest as a result of this procedure were subsequently ignored.

For each sampling location, i.e. PSU, determined by the procedure just described, the CNI collected data at each of a series of points within that PSU. In order to determine the positions of these sampling points, an aerial photograph of the sampling location was obtained. A spinner or template consisting of a grid of small holes was then centered over the photograph and spun. A deterministic procedure was used to choose a hole for the location of the spinner in a fashion that allowed some variety in the choice of the spinner location without introducing personal bias.[2] When the template came to rest, the location of each hole was marked on the photograph. The first point in the upper left

---

[2] The procedure for selecting the spinner hole is described in Appendix #2 of the National Handbook for Updating the Conservation Needs Inventory (U.S.D.A., Washington, D.C., August 1966).

Table 1.1. Sampling Rates (%) Which Provide Standard Relative Precision of County Level Estimates for 10 Size-classes and 3 Sizes of Unit

County size-class	(Square miles)	Size of unit (PSU) 40 acres	160 acres	640 acres
1	47 and less	16	32	64
2	48 - 95	8	16	32
3	96 - 191	4	8	16
4	192 - 383	2	4	8
5	384 - 767	1	2	4
6	768 - 1,535	1/2	1	2
7	1,536 - 3,071	1/4	1/2	1
8	3,072 - 6,143	1/8	1/4	1/2
9	6,144 - 12,287	1/16	1/8	1/4
10	12,288 and over	1/32	1/16	1/8

*Source: Taylor, Howard L. Statistical Sampling for Soil Mapping Surveys, June 1962, courtesy of the Iowa State University Statistical Laboratory.

corner of the sample site was point number one. The points were then numbered consecutively along a line proceeding from left to right and/or up. The consecutive numbering of the sampling points then continued in the same manner with the line of points just below the first line. This procedure continued until all points in the sample area had been numbered as illustrated in Figure 2. These points constitute an aligned two-dimensional systematic sample within each selected PSU.[3] Such an alignment of points in a strictly North-South and East-West manner should be avoided because of the tendency to develop land use in such a pattern; the spinning of the template alleviates this.

Various sampling templates were prepared so that template and aerial photograph scales could be matched to obtain a constant sampling density. It was most convenient to assign the sampling points in local USDA offices, since local Soil Conservation Service offices generally had the needed aerial photographs in their files. However, the local USDA personnel did not always follow the sampling protocol specified by the design. For instance, it appears that the templates were not spun for Nevada, and the template was often not properly matched to the photograph scale in New Mexico.

Exhibit 1 is a photocopy of an aerial photograph of a specific 160-acre CNI sample site with 34 consecutively numbered sampling points. The point density of the template used for this site was the standard point density intended for all sites, except for the 640-acre sites. The point density of the templates used for 640-acre sites was one-fourth that of the other sites, since 640-acre sites were used only in homogeneous land areas. Thus, 160-acre and 640-acre sites usually received from 34 to 39 sampling points and 40-acre sites usually received between 9 and 11 points.

Exhibit 2 is a photocopy of the data collection form used to record the data for the 34 sampling points shown in Exhibit 1. The data items that were used in determining the Rural Soils Network (RSN) subsample were the Field Mapping Symbols and the Land Use Codes. In particular, this information was used to classify each sampling point as either a cropland point or a noncropland point as shown in Table 1.2. It should be noted that sampling points that inadvertently fell into areas outside the target population, i.e., urban areas, water areas, and federal noncropland, were classified as noncropland points.

The counts of cropland and noncropland points were accumulated as shown in Exhibit 3 for the purpose of selecting the RSN subsample from the CNI sample. The data for the CNI sites shown in Exhibits 1 and 2 appear on the fourth line in Exhibit 3. In particular, Exhibits 1 and 2 are for State 16, Kansas; County 66, Nemaha; site number 5-2-2R. A total of 34 points were sampled at this site and 19 of these points were designated as cropland points. Thus, the proportion of cropland points at this site was $19/34 \doteq 0.55882$. However, the sampling rate in Nemaha County was 2.257%; i.e., the ratio of sampled acreage to total inventory acreage in Nemaha County was about 0.02257. Thus, in order to adjust the cropland proportion to a standard 2% sampling rate, the "cropland ratio" was computed as

---

[3] See, for example, Cochran, W. G. [1977, pg 228]. Sampling Techniques. Wiley, New York.

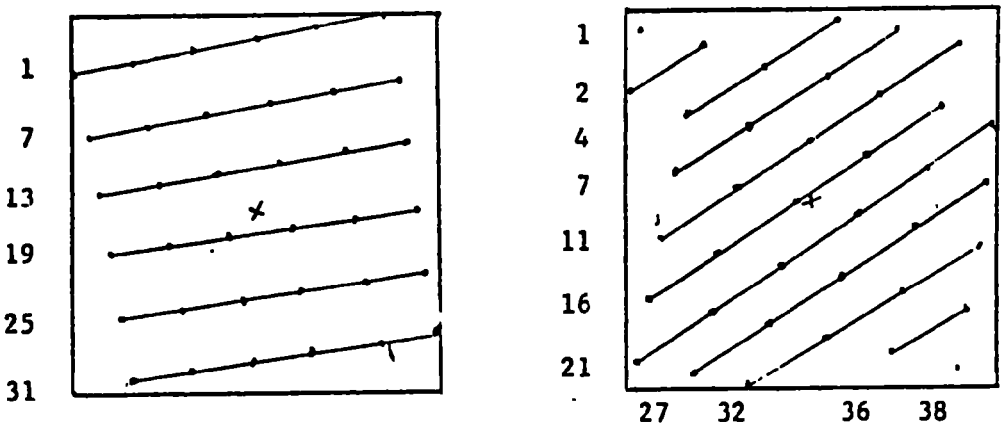


Figure 2: Sample Points on a 160-acre Sample Area

Note: The numbers above are the point numbers for the first points on each line.

Source: Appendix #2 of the National Handbook for Updating the Conservation Needs Inventory (U.S.D.A., Washington, D.C., August 1966).

Table 1.2. Dichotomization of the Land Use Code*

CROPLAND CATEGORIES

Land Use Codes		
Nonirrigated	Irrigated	
L10	L11	Corn and sorghums
L20	L21	All other row crops
L30	L31	Close grown field crops
L40		Cultivated summer fallow
L50	L51	Rotation hay and pasture
L60	L61	Hayland
L90	L91	Orchards, vineyards, and bush fruits

NONCROPLAND CATEGORIES

Land Use Codes		
Nonirrigated	Irrigated	
L70		Conservation use only
L80		Temporarily idle cropland
L00		Open land formerly used for crops
P10	P11	Pasture
P20		Range
F10		Commercial forest
F20		Noncommercial forest
H10		Other land in farms
H20		Other land not in farms

Field Mapping Sybmol

UB	Urban or built-up area
FED	Federal noncropland
W1	Water area of more than 40 acres
W2	Water area of 40 acres or less
W3	Intermittent water area

*Source: Memorandum entitled "Soil Monitoring Program--Sampling Design" from Leo G.K. Iverson to USDA PPC Inspectors.

CONS. NEEDS SAMPLE NO. 5-2-2R

State KANS County NEMAHA

Photo No. YX-2CC-171

Ownership

Area

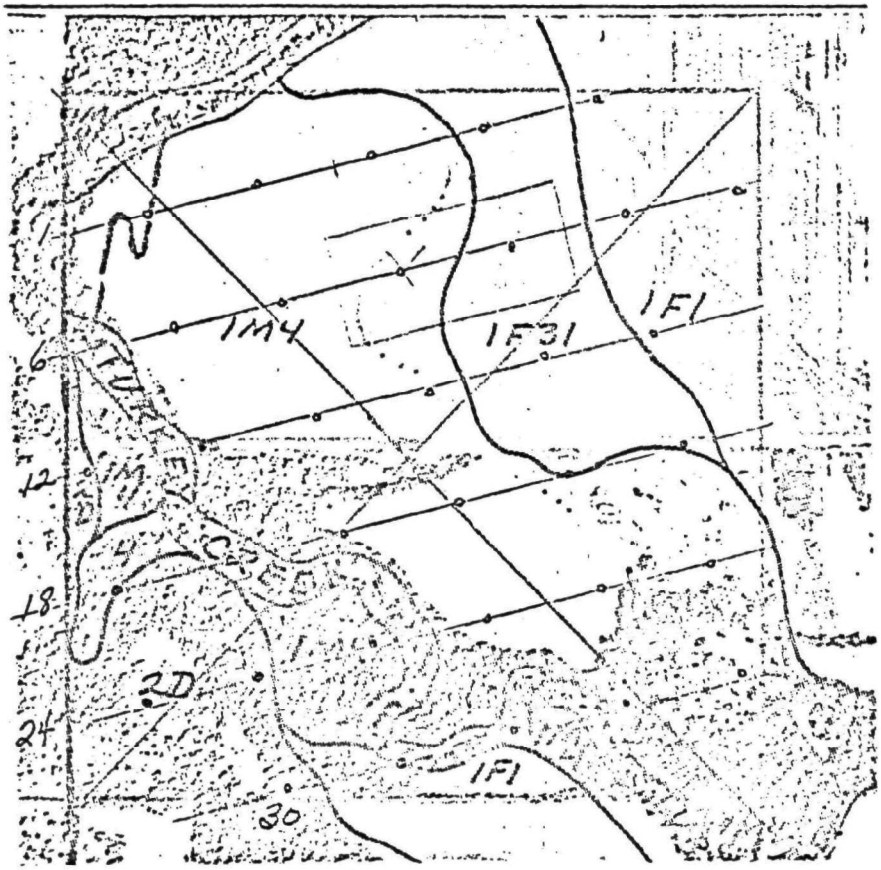


Exhibit 1: Aerial Photograph of a CNI Sample Location

*Source: Sampling files maintained by the EPA Field Studies Branch, Washington, D.C.

SCS-263
10-65

SAMPLE AREA RECORD

Soil, Land Use and Treatment Needs

SOIL CONSERVATION SERVICE

State _____ County _____

State _____ County _____ SCD _____ Sample Area _____

Size _____ Ownership _____ Land Res. Area _____ Land Res. Reg. _____ Agr. Sub. Reg. _____

Sub Basin _____ Technician _____ Date 10-31-66

Line or Pt.	Field Mapping Symbol	State Code: Soil					State Code: Slope Class				State Code: Eros. Cl.		Land Capability Unit					Land Use			Trt. Nds.
31 32		50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
1	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1
2	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1
3	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1
4	1F31	4	3	7	H		A		0	1	0		1	—	—	—	H	L	1	0	1
5	1F1	4	0	5	A		A		0	1	0		1	—	0	5	H	L	1	0	1
6	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1
7	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1
8	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1
9	1F31	4	3	7	A		A		0	1	0		1	—	—	—	H	L	1	0	1
10	1F1	4	0	5	A		A		0	1	0		1	—	0	5	H	L	1	0	1
11	1F1	4	1	5	A		A		0	1	0		1	—	0	5	H	L	1	0	1
12	—D	2	5	7	A		E		0	0	0		6	—	0	7	H	E	2	1	X
13	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	M	1	0	X
14	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1
15	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1
16	1F31	4	3	7	A		A		0	1	0		1	—	—	—	H	L	1	0	1
17	1F1	4	0	5	A		A		0	1	0		1	—	0	5	H	L	1	0	1
18	2D	2	5	7	A		E		0	0	0		6	—	0	7	H	E	2	1	X
19	W2																				X
20	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	
21	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1 X
22	1F31	4	3	7	A		A		0	1	0		1	—	—	—	H	L	1	0	X
23	1F31	4	3	7	A		A		0	1	0		1	—	—	—	H	L	1	0	X
24	2D	2	5	7	A		E		0	0	0		6	—	0	7	H	E	2	1	X
25	2D	2	5	7	A		E		0	0	0		6	—	0	7	H	E	2	1	X
26	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	X
27	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	1
28	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	
29	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	
30	2D	2	5	7	A		E		0	0	0		6	—	0	7	H	E	2	1	X
31	W2																				X
32	W2																				X
33	1M4		1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	X
34	1M4	2	1	1	A		A		0	1	0		1	—	0	5	H	L	1	0	X
35																					
36																					
37																					
38																					

Exhibit 2: CNI Data Collection Sheet

*Source: Sampling files maintained by the EPA Field Studies Branch, Washington, D.C.

Exhibit 3: Accumulation Used in Selecting the Pesticide Residue Network Subsample

State	County	LCNI No.	Total Points	Crop Points	Cropland Ratio	Noncropland Ratio	Cropland Accumulation	Noncropland Accumulation
16	66	5 1 2R	33	18	.48334	.40278	1988.58597	2052.83459
16	66	5 1 3R	32	5	.13845	.74767	1988.72443	2053.58226
16	66	5 2 1R	33	10	.26852	.61760	1988.99295	2054.19987
16	66	5 2 2R	34	19	.49519	.39093	1989.48814	2054.59081
16	66	5 2 3R	36	22	.54152	.34460	1990.02967	2054.93541
16	66	5 3 1R	24	21	.77536	.11076	1990.80503	2055.04618
16	66	5 3 2R	36	31	.76305	.12307	1991.56809	2055.16925
16	66	5 3 3R	35	20	.50635	.37977	1992.07445	2055.54902
16	66	5 4 1R	35	24	.60763	.27849	1992.68208	2055.82752
16	66	5 4 2R	34	32	.83400	.05212	1993.51608	2055.87965
16	66	5 4 3R	33	24	.64445	.24167	1994.16054	2056.12132
16	67	1 1 1R	35	0	0.00000	.97847	1994.16054	2057.09979
16	67	1 1 2R	35	21	.58708	.39138	1994.74762	2057.49118
16	67	1 1 3R	36	11	.29897	.67949	1995.04660	2058.17067
16	67	1 2 1R	32	10	.30577	.67269	1995.35237	2058.84337
16	67	1 2 2R	36	23	.62513	.35333	1995.97751	2059.19670
16	67	1 2 3R	37	0	0.00000	.97847	1995.97751	2060.17517
16	67	1 3 1R	35	30	.83868	.13978	1996.81619	2060.31496
16	67	1 3 2R	36	5	.13589	.84257	1996.95209	2061.15753
16	67	1 3 3R	37	15	.39667	.58179	1997.34877	2061.73932
16	67	1 4 1R	35	2	.05591	.92255	1997.40468	2062.66188
16	67	1 4 2R	40	13	.31800	.66046	1997.72269	2063.32234
16	67	1 4 3R	32	24	.73385	.24461	1998.45654	2063.56696
16	67	2 1 1R	36	15	.40769	.57077	1998.86423	2064.13774
16	67	2 1 2R	37	32	.84624	.13222	1999.71048	2064.26996
16	67	2 1 3R	38	16	.41198	.56648	2000.12247	2064.83644
16	67	2 2 1R	34	10	.28778	.69068	2000.41025	2065.52713
16	67	2 2 2R	36	10	.27179	.70667	2000.68205	2066.23380
16	67	2 2 3R	33	22	.65231	.32615	2001.33436	2066.55996
16	67	2 3 1R	35	33	.92255	.05591	2002.25692	2066.61587
16	67	2 3 2R	38	23	.59223	.38623	2002.84915	2067.00211
16	67	2 3 3R	38	8	.20599	.77247	2003.05514	2067.77459
16	67	2 4 1R	35	17	.47525	.50321	2003.53040	2068.27780
16	67	2 4 2R	37	17	.44956	.52890	2003.97997	2068.80670
16	67	2 4 3R	38	10	.25749	.72097	2004.23746	2069.52768
16	67	3 1 1R	35	19	.53116	.44730	2004.76863	2069.97498
16	67	3 1 2R	35	25	.69890	.27956	2005.46754	2070.25454
16	67	3 1 3R	36	11	.29897	.67949	2005.76651	2070.93404
16	67	3 2 1R	37	20	.52890	.44956	2006.29542	2071.38360
16	67	3 2 2R	37	12	.31734	.66112	2006.61276	2072.04473
16	67	3 2 3R	37	10	.26445	.71401	2006.87721	2072.75875
16	67	3 3 1R	38	10	.25749	.72097	2007.13470	2073.47973
16	67	3 3 2R	36	22	.59795	.38051	2007.73266	2073.86024
16	67	3 3 3R	41	9	.21478	.76368	2007.94744	2074.62393
16	67	3 4 1R	38	38	.97847	0.00000	2008.92591	2074.62393

Source: Sampling files maintained by the EPA Field Studies Branch, Washington, DC

$$\frac{19}{34} \cdot \frac{2}{2.257} \doteq 0.49519$$

This cropland ratio of 0.49519 was then added to the cropland accumulation, which was the sum of the cropland ratios for all previously listed sites in the State.

The procedure used to obtain the noncropland accumulation was identical to that just described for the cropland accumulation. However, it was considered desirable to include federal noncroplands in the RSN noncropland sample. Although federal noncroplands had not been mapped by the CNI, the CNI sampling procedure did assign PSU's in federal noncropland areas. That is, the CNI sample sites were selected without regard to federal land status. Whenever a CNI sample site fell entirely in a federal noncropland area, no CNI sampling points were assigned to the site. In order to obtain coverage of these federal noncropland sites by the RSN noncropland sample, the sampling staff obtained a list of the CNI sample sites in federal noncropland areas for each State. The sampling staff then inserted a "dummy" CNI record into the listings of the type shown in Exhibit 3 for each CNI site that fell entirely in federal noncropland. Each dummy record showed zero cropland points, and a total number of points appropriate for the size of the sample site, e.g., 36 points for a 160-acre PSU.

The grand total of the cropland accumulation from Kansas was 3426.67927, and for noncropland it was 3131.41689. The total of these accumulations, 6558.09616, was employed for estimation of the proportion of cropland and noncropland acreage in Kansas. In particular, the estimate of the proportion of cropland acreage in Kansas was

$$\frac{3426.67927}{6558.09616} = 52.25112878\%$$

This procedure provides a direct estimate of the proportion of cropland acreage in the State. This estimated proportion of cropland was multiplied by an estimate of the total land area in Kansas, namely 52,510,720 acres, to yield an estimated cropland acreage in Kansas of

$$(.5225112878)\ (52{,}510{,}720) = 27{,}437{,}444 \text{ acres.}$$

This same procedure was used for all States.

1.2.2.2 The RSN Survey

The Rural Soils Network (RSN) selected two subsamples from the CNI sample sites, a cropland sample and a noncropland sample. The sample design of the RSN specified that the subsamples would contain 0.025 percent of the cropland acreage and 0.0025 percent of the noncropland acreage in each State. Thus the cropland sample in Kansas was to consist of

$$(0.00025)\ (27{,}437{,}444) \quad = \quad 6{,}859.36 \text{ acres.}$$

Each RSN sample site was to be a 10-acre plot with an equal number of plots sampled in each of four years. Thus, the number of cropland sample sites to be selected in Kansas in each of four years of sampling was

$$\frac{6{,}859.36 \text{ acres}}{(10 \text{ acres/site})\ (4 \text{ years})} \quad = \quad 171 \text{ sites/year.}$$

The number of 10-acre noncropland sites to be sampled in each State was determined in exactly the same manner. Each RSN site was to be sampled a second time four years after the initial sampling to determine rates of change in pesticide residues. Implementation of this design for all States resulted in the sample sizes shown in Table 1.3. This sample was expected to yield reasonably precise estimates for cropping regions and some of the larger States.

Having determined the number of RSN cropland sites to be selected in a State, a systematic subsample of CNI cropland points was selected from the cropland accumulation for the State. Each CNI cropland point selected was used to locate a 10-acre RSN cropland sample site. It is easiest to explain this procedure by example. The total of the cropland accumulation for Kansas was 3426.67927, and 171 cropland sites were to be surveyed in each of 4 years. Thus, the starting point for the sample in Kansas was a random number between zero and

$$\frac{3426.67927}{(4)\ (171)} \quad \doteq \quad 5.00976$$

The random number chosen was 0.27889, which determined the selection of the first RSN cropland site. All RSN cropland sites in Kansas then resulted from a sequence number of the form

$$0.27889 + k\ (5.00976) \text{ for } k = 0,1,2,\ldots, [(4)(171)-1 = 683]$$

The RSN cropland site in Kansas that was considered previously in this discussion resulted from the sequence number

$$0.27889 + (397)\ (5.00976) \quad = \quad 1989.15361,$$

as seen on the first line of Exhibit 4.

The sequence number 1989.15361 not only determined that CNI site 5-2-2R of Nemaha County, Kansas was to be included in the cropland sample of the RSN; it also specified that a particular point at this site was to be used to locate the 10-acre RSN site. Hence, one of the 19 cropland points at this site was determined by interpolation. From Exhibit 3, the following cropland accumulations were obtained for interpolation:

State	County	CNI Site	Cropland Accumulation
16	66	5-2-1R	1988.99295
16	66	5-2-2R	1989.48814

Table 1.3: Design Sample sizes for the Rural Soils Network*

Census Division State	Component		Total
	Cropland	Noncropland	
New England	80	96	176
Maine	32	48	80
New Hampshire	8	12	20
Vermont	20	12	32
Massachusetts	8	12	20
Rhode Island	4	4	8
Connecticut	8	8	16
Middle Atlantic	320	128	448
New York	152	60	212
New Jersey	20	12	32
Pennsylvania	148	56	204
East-North Central	1648	224	1872
Ohio	276	36	312
Indiana	312	28	340
Illinois	568	32	600
Michigan	220	68	288
Wisconsin	272	60	332
Pacific	600	456	1056
Washington	180	92	272
Oregon	152	140	292
California	268	224	492
West-North Central	3596	456	4052
Minnesota	488	80	568
Iowa	608	28	636
Missouri	328	76	404
N. Dakota	636	48	684
S. Dakota	424	80	504
Nebraska	428	80	508
Kansas	684	64	748

(continued)

*Source: Wiersma, G.B., Sand, P.F., and Cox, E.L. (1971). A sampling Design to Determine Pesticide Residue Levels in Soils of the Conterminous United States. Pesticides Monitoring Journal 5(1), pp. 63-66.

Table 1.3: Design Sample Sizes for the Rural Soils Network (continued)

Census Division / State	Component		Total
	Cropland	Noncropland	
South Atlantic	556	376	932
Delaware	12	4	16
Maryland	52	12	64
Virginia	84	56	140
W. Virginia	24	36	60
N. Carolina	124	68	192
S. Carolina	68	40	108
Georgia	120	80	200
Florida	72	80	152
East-South Central	452	244	696
Kentucky	124	52	176
Tennessee	112	56	168
Alabama	92	72	154
Mississippi	124	64	188
West-South Central	1300	552	1852
Arkansas	188	64	252
Louisiana	108	60	168
Oklahoma	260	84	344
Texas	744	344	1088
Mountain	916	1280	2216
Montana	340	200	540
Idaho	132	120	252
Wyoming	68	148	216
Colorado	240	140	380
New Mexico	40	192	232
Arizona	36	176	212
Utah	48	128	176
Nevada	12	176	188
Grand Total	9468	3812	13280

Exhibit 4: RPRN Cropland Sampling

State	County	CNI No			Total Points	Crop Points	Cropland Accumulation	Noncropland Accumulation	Sequence Number	Sample Point	Sample Year	Sample Type
16	66	5	2	2R	34	19	1989.48814	2054.59081	1989.15361	7	2	CROP
16	67	1	1	2R	35	21	1994.74762	2057.49118	1994.16337	1	4	CROP
16	67	2	1	2R	37	32	1999.71048	2064.26996	1999.17313	12	3	CROP
16	67	2	4	3R	38	10	2004.23746	2069.52768	2004.18289	8	1	CROP
16	67	3	4	3R	37	16	2009.60064	2075.90614	2009.19265	1	2	CROP
16	67	4	4	3R	34	25	2014.88694	2082.36148	2014.20241	2	4	CROP
16	68	1	3	1R	35	35	2019.44149	2085.01413	2019.21217	31	3	CROP
16	68	1	5	2R	34	22	2024.86161	2086.80121	2024.22193	10	1	CROP
16	68	2	3	1R	35	20	2029.61132	2091.06050	2029.23169	13	2	CROP
16	68	2	4	2R	33	33	2034.96667	2092.91235	2034.24145	20	4	CROP
16	68	3	1	2R	33	30	2040.60010	2094.48612	2039.25121	6	3	CROP
16	68	3	4	1R	35	35	2044.89868	2095.59294	2044.26097	23	1	CROP
16	68	3	6	1R	36	36	2049.73780	2096.15922	2049.27073	27	2	CROP
16	68	4	1	2R	36	36	2054.44250	2096.85992	2054.28049	33	4	CROP
16	68	4	5	2R	35	35	2060.84229	2101.27093	2059.29025	5	3	CROP
16	68	5	1	1R	32	32	2064.44589	2101.27093	2064.30001	30	1	CROP
16	68	5	4	1R	14	14	2070.56419.	2104.16163	2069.30977	5	2	CROP
16	68	5	6	2R	35	35	2075.41231	2108.32251	2074.31953	14	4	CROP
16	69	1	5	1R	34	32	2080.68472	2116.01305	2079.32929	8	3	CROP
16	69	2	2	2R	33	16	2084.46431	2119.64086	2084.33905	14	1	CROP
16	69	3	1	1R	36	36	2091.14824	2122.21618	2089.34881	2	2	CROP
16	69	3	3	1R	36	31	2094.48451	2124.43546	2094.35857	29	4	CROP
16	69	4	3	1R	34	27	2100.71786	2131.16506	2099.36853	3	3	CROP
16	69	5	1	2R	34	34	2106.11503	2135.02714	2104.37809	3	1	CROP
16	69	5	4	2R	34	24	2110.07084	2140.33058	2109.38785	12	2	CROP
16	70	1	1	3R	34	11	2114.63683	2142.35288	2114.39761	3	4	CROP
16	70	2	4	3R	34	18	2119.49500	2157.68684	2119.40737	15	3	CROP
16	70	4	2	2R	33	30	2124.58688	2168.94097	2124.41713	25	1	CROP
16	70	5	2	2R	35	18	2129.83954	2175.22667	2129.42689	3	2	CROP
16	71	1	4	2R	34	23	2135.65680	2189.03146	2134.43665	1	4	CROP
16	71	3	1	2R	31	12	2139.79961	2199.62161	2139.44641	7	3	CROP
16	71	4	1	1R	35	27	2145.38178	2210.61402	2144.45617	10	1	CROP
16	71	4	4	1R	16	15	2150.09456	2215.10934	2149.46593	10	2	CROP
16	71	4	5	2R	32	32	2155.61942	2215.10934	2154.47569	13	4	CROP
16	71	5	3	1R	35	17	2159.81989	2220.11697	2159.48545	11	3	CROP
16	72	1	2	2R	35	31	2164.69412	2223.66753	2164.49521	24	1	CROP
16	72	2	2	1R	36	27	2170.11112	2232.47518	2169.50497	4	2	CROP
16	72	2	5	2R	36	6	2174.54683	2237.52257	2174.51473	5	4	CROP
16	72	3	3	1R	33	8	2179.64336	2240.01252	2179.52449	4	3	CROP
16	72	4	3	1R	32	21	2184.83662	2249.04391	2184.53425	11	1	CROP
16	73	1	4	3R	34	34	2190.16934	2256.82757	2189.54401	11	2	CROP
16	73	1	6	3R	40	34	2195.15867	2257.36814	2194.55377	8	4	CROP
16	73	2	3	1R	41	39	2200.14425	2258.83411	2199.56353	14	3	CROP
16	73	2	5	2R	33	32	2204.75086	2260.67905	2204.57329	26	1	CROP
16	73	3	1	2R	36	25	2209.84577	2261.11404	2209.58305	15	2	CROP

Source: Sampling files maintained by the EPA Field Studies Branch, Washington, D.C.

The interpolation proceeded as follows:

$$\frac{1989.15361 - 1988.99295}{1989.48814 - 1988.99295} \cdot (19) = 6.16 \rightarrow 7.$$

The interpolation figure was rounded up since an integer from 1 to 19 was required. In this case, the seventh cropland point at CNI site 5-2-2R was to be used to locate the RSN cropland site, as is also specified in Exhibit 4.

Once a defining point for a RSN site had been selected, an adjacent second cropland (or noncropland) point was required in order to completely determine the location of the 10-acre RSN site. If X is used to denote the defining cropland point selected from the CNI sample, an adjacent cropland point was to be determined by considering the other CNI sample points in the order indicated below :

	4	
3	X	1
	2	

If an acceptable second cropland point could not be located as indicated, then the next cropland point in the listing was taken as a first point and the routine repeated.[4] This procedure was implemented in the USDA offices prior to field work, and some discretion was allowed. The intention was clearly that an RSN cropland site should not be placed at an isolated cropland point.

After two points had been selected, a designation was made on an aerial photograph or other map of a 10-acre site with these points centrally located. Attention was given to making the boundaries conform with natural physical features as much as possible. Implementation of this procedure can be illustrated by Exhibits 1 and 2. The design specified that the seventh cropland point was to be used to locate the RSN site. From Exhibit 2, it can be seen that the seventh cropland point is the eighth CNI sample point. In Exhibit 1, it can be seen that the depicted RSN site was, indeed, centered about the eighth and ninth CNI sample points, both cropland points.

The field person was permitted to adjust the boundaries of the designated 10-acre RSN site and was expected to prepare records so that the site could be readily relocated for subsequent sampling at 4-year intervals. The final sample location was to be not less than 8 acres. If the designated site should prove to be totally unacceptable,[5] the field person was permitted the following alternatives in order of preference:

1) Try to find 10 acres within the CNI site that are acceptable.

2) Try to find 10 acres within one-fourth mile of the CNI site that are acceptable.

---

[4] Memorandum entitled "Soil Monitoring Program -- sampling design" from Leo G.K. Iverson to USDA PPC Inspectors.

[5] The authors were not able to find an explicit definition of "totally unacceptable."

3) Try to locate two smaller sites within the CNI site that equal 10 or nearly 10 acres. Sample as if they were a single site.

4) Request the USDA staff at Hyattsville to re-select the CNI site.[6]

Substitute CNI sites were selected in a number of cases. The substitutes were chosen from within the same county as the original site. An effort was made to choose a substitute CNI site with approximately the same proportion of cropland points as the original CNI site. However, since a random sequence number was not used to determine the substitute site, it was necessary to randomly designate a point within the substitute CNI site to locate the 10-acre RSN site. It is not clear that this randomization was always performed.

There are several reasons why substitute CNI sites were sometimes required. Re-selections were performed by the USDA staff at Hyattsville before the sample went to the field when the selected CNI site was already in use by the USDA. For example, the selected CNI sites were occasionally found to be in use by

a) the Soil Conservation Service for their crop estimates,
b) the Economic Research Service for their Pesticide Use Survey,
c) the June Enumerative Survey of the Statistical Reporting Service.

Re-selections were sometimes necessary after the sample went to the field because the land owner refused to cooperate. Some re-selection was necessary because of a change of land use status. Unfortunately, substitute sites are not designated as such on the computer records. This is especially problematic if a substitute was selected in the second round of data collection. First round and second round data cannot be compared directly for a site if a substitute has been used.

1.2.3 Limitations as a Monitoring Network

The Rural Soils Network (RSN) design specified that 0.025 percent of the cropland acreage and 0.0025 percent of the noncropland acreage was to be sampled in each State. This criterion resulted in sample sizes that vary considerably from one State to another. Rhode Island received the fewest sampling units, four each of cropland and noncropland. Texas received the most, 744 cropland sites and 344 noncropland sites. Thus, reliable estimates of average pesticide levels are not available for some geographic areas. This is a minor limitation because estimates are not generally required for small geographic areas. The deletion of some States when the design was implemented restricts the population to which inferences are valid, however.

---

[6] Shepherd, D.R. PPC Division Memorandum 804.3 concerning "Guidelines for collecting sample for the National Soil Monitoring Program--1969."

More significantly, the following factors must be noted:

- The current design was found to be too expensive to operate.
- The network as it stands was not designed to monitor non-pesticide toxic materials, hence may be inadequate particularly for non-agricultural areas and localized contaminants.
- The stratification is now 15 years out-of-date, which means losses in efficiency.
- The two phase design renders estimating precision difficult.

1.2.4 Uses in Regulatory Action

The Rural Soils Network (RSN) could be used to identify pesticides and other widely dispersed toxic substances for which regulatory action is desirable. Each sample site of the RSN was to be sampled every four years. Thus, significant increases in average levels of specific substance, could potentially be discovered. Moreover, since residue levels were determined for both soils and crops, the relationship between soil and crop residue levels could be used to identify potentially dangerous levels of soil residue. For example, if a pesticide level in corn that is dangerous for humans has been identified, the relationship between soil and corn concentration of that pesticide could be used to determine a corresponding dangerous level of the pesticide in soil.

The RSN could also be used to monitor the effects of regulation of specific toxic substances. Because each RSN site is sampled every four years, the network could monitor the effect of the regulation on levels of the toxic substances in soils and crops.

The RSN may be of limited use, however, in identifying specific violators of regulatory action. This situation results from the very design of the RSN. The RSN is designed to be sites selected by a random process at a given sampling rate with the location of specific sites being confidential to protect the farm operator. Specific localities of interest may not enter the RSN sample, but the design framework could serve as the basis for special studies in suspected "hot spots."

1.2.5 User Needs and Historical Uses of the Data

The historical objectives of the Rural Soils Network (RSN) were as follows:[7]

(1) Determine levels of pesticides and other pollutants in the agricultural environment.

(2) Observe trends in pollutant levels through time.

(3) Determine the degree to which crops are contaminated.

(4) Determine the levels of various pollutants in agricultural waters.

---

[7] Shepherd, D. R. PPC Division Memorandum 804.3 concerning "Guidelines for collecting samples for the National Soil Monitoring Program--1969."

(5) Determine the concentration of certain pollutants at various depths in the soil profile.

(6) Review program findings with recommendation of appropriate actions in mind.

The six objectives listed above comprise the major historical user needs for the RSN data. The regulatory uses considered in section 1.2.4 are included in objective (6) above.

The implementation of the RSN allows only partial fulfillment of the six objectives listed above. It appears that objective (5) has been abandoned since soil data has been collected only for the top three inches of soil. Objective (4) has only been partially addressed by sampling pond water and sediment during a single fiscal year. Most States have follow-up data with which to address objective (2) for only one-fourth of the cropland sites and none of the noncropland sites.

1.3. Alternate Survey Designs for the RSN

The Rural Soils Network (RSN) is a probability sample of the rural areas of the conterminous United States. A probability sample is essential as an objective basis for making inferences. The RSN is, however, a subsample of the 1967 Conservation Needs Inventory (CNI). It relies upon the CNI to identify the cropland and noncropland strata, as in double sampling schemes. As the 1967 CNI became outdated, sites were found in the field to no longer belong to the intended stratum, cropland or noncropland. It has been the practice for the field personnel of the RSN to use substitute sites in these cases. The use of substitute sites tends to destroy the probabilistic nature of the sample and is not generally recommended, however.[8] Resumption of RSN data collection is likely to result in many sites being misclassified.

Thus, sampling considerations alone suggest that a new RSN sample is needed. In addition, a new sampling design should address the problem of monitoring toxic substances other than agricultural chemicals and should attempt to reduce the cost of the monitoring network. The expense of the RSN led to purposive deletion of entire States in the past, which restricts the population to which valid inferences can be made. Various alternative designs will now be considered.

1.3.1 Design Option One

A minimal change alternative would be to subsample the current RSN on a probability basis. This option mainly addresses the problem of the cost of the RSN, however it does also address the need for regional and national estimates. Any need to eliminate reliance upon the 1967 CNI is not addressed.

This option does have some advantages, however. It's main advantage is that it can be implemented quickly and easily, possibly while other alternatives are under development. Another advantage is that direct comparison could be made to the data collected from 1968 to 1975.

---

[8] See, e.g., page 386 of Kish, Leslie (1965). Survey Sampling. Wiley.

Careful treatment of the sites found to no longer belong to the intended stratum would be necessary. There are at least three ways that these sites could be handled. One possibility would be to drop these sample sites entirely. There would be a loss in precision for estimates, and the sampling weights would have to be adjusted to reduce the bias that would result from deletion of these sites. Alternatively, substitute sites could be selected, as has been done historically with this sample. However, the use of substitute sites introduces bias that cannot be measured or adjusted. Finally, sites can be retained as selected. This keeps the initial weight correct and provides unbiased estimates at the cost of a decrease in precision. The computerized data records would need to indicate the resolution of each of these cases, whether they were all dropped, or substitutes were selected, or retained in their original strata. If as many as 10 percent of the sample sites require either deletion or substitution, this design option may not be reasonably efficient..

Data analysis problems would be aggravated by subsampling the present RSN. The deep stratification of the 1967 CNI results in stratification benefits for sample variances for the RSN. However, the sparseness of the RSN sample in comparison to the CNI sample makes recovery of the stratification effects difficult (See section 1.7). The major problem is that many counties have no more than one RSN site. The magnitude of this problem would necessarily increase with a subsample of the current RSN.

Thus, replicate subsamples are recommended if this option is to be implemented, even if it is only on a temporary basis. For example, if 50 percent of the RSN sites are to be surveyed, five subsamples that each comprise a 10 percent subsample could be used. At least five replicate subsamples should be selected. A defensible procedure for selecting the replicate subsamples would be to first order the RSN sites by States and CNI strata within States, then independent systematic subsamples could be selected. This procedure would insure representation of all states and as much CNI stratification as possible in each of the replicate subsamples (or technically 'pseudo-replicate' subsamples).

The use of replicate subsamples would make it possible to estimate easily sample variances by using the theory of replicate subsamples.[9] The results of interest would initially be tabulated separately for each independent subsample. The variance of these results treated as independent measurements provides a simple, unbiased estimate. The resulting variance estimate captures all design effects, although stratification effects and design effects are not separately estimable. This is not of major consequence for the present RSN sample, since only one stage of sampling is employed within CNI strata.

It might also be useful to select the subsamples at different rates within domains of interest. The present RSN sample has widely different sample sizes within the Census Divisions, and within the cropping regions. If cropping regions comprise the major domains of interest, they could be subsampled at differential rates so that each received about the same

---

[9] See, e.g., page 19 of Cochran, W.G., Mosteller, F., and Tukey, J.W. [1975]. Principles of Sampling. Journal of the American Statistical Association, 70: 13-35.

number of RSN sites. Alternatively, Census Divisions could be subsampled at differential rates, which might considerably reduce the sample size in some of the larger States, like Texas.

Finally, identification of strata of special interest within the domains just considered, could be used to increase the possibility of finding toxic substance residues. For example, the noncropland RSN sites could be stratified into industrial and nonindustrial areas. Sites in nonindustrial areas could then be sampled at a lower rate than sites in industrial areas. Stratification according to whether or not toxic residues have previously been found at the site may be useful also. Widely different sampling rates would not be used for these strata, however, because they would form a far from homogeneous group.

1.3.2 Design Option Two

The present RSN sample is a subsample of the 1967 Conservation Needs Inventory (CNI). A design analogous to the design that produced the present RSN sample could be based upon the 1982 National Resources Inventory (NRI). Use of the 1982 NRI would provide up-to-date land use information. The NRI was designed by the Statistical Laboratory at Iowa State University, and is currently being conducted by the Soil Conservation Service. The design of the NRI is similar to that of the CNI, except that the standard sampling practice is to collect land use data for exactly three random sampling points within each primary sampling unit (PSU) of the NRI.[10] Also, the NRI is based upon a more dense sample than was the CNI. Consequently, data collection for the NRI is over three years, 1980 to 1982.

The procedure used to select the RSN subsample from the CNI sampling points resulted in a sample that was essentially self-weighting within States where only one size of PSU was used (See Appendix D). Equal weighting was an important consideration before the development of computer software for the analysis of unequal probability samples. The unweighted analysis of data from sample sites selected with unequal probabilities can well lead to spurious conclusions.

Since software is now available for the analysis of unequal probability samples, an improved subsampling procedure can be devised. The goal of the subsampling procedure is to obtain adequate precision at minimum cost. This can be accomplished by identifying areas where toxic residues are likely to be found and giving these areas a higher probability of selection. It is, of course, important that all areas have a positive probability of being in the sample so that statistical inferences will be valid for the entire population.

It is suggested that counties be used as primary sampling units for the second phase sample. The data from the present RSN suggests that counties are generally rather heterogeneous with respect to toxic residues. Thus, it would be advantageous to select relatively few counties with a relatively large number of sample sites, say 5 to 10, within each sample county. The use of counties as PSU's will reduce

[10]The NRI sampling design also includes pilot studies of alternative sampling designs in California, Louisiana, and Maine. In Louisiana (and in 40-acre PSU's), there is only one random sampling point within a PSU.

travel costs associated with data collection. More importantly, however, smaller areas like counties can be stratified more effectively into areas where toxic residues are likely to be found.

The RSN sample sites are to be located at NRI sample points. Thus, sample counties are selected from the counties occuring in the NRI sample, and so that counties where toxic substance residues are likely to occur have a greater chance of selection. Thus, it is suggested that counties be selected with probability proportional to size (PPS), where the size measure is a measure of the likelihood for finding toxic residues. Selection of PSU's with PPS sampling is a common technique with resulting variances of estimates reduced to the extent that the size measure is correlated with items of interest. Variables that can be used to construct county size measures include:

(1) Proportion of county acreage in cropland.
(2) Proportion of county acreage in heavy industry.
(3) Intensity of agricultural activity.
(4) Degree of industrialization.
(5) Predominant crops.
(6) Predominant industries.
(7) Predominant soil types.
(8) Climate

Counties should be selected with PPS sampling within Census Divisions, cropping regions, or some other domains to insure adequate representation of the major domains of interest.

After sample counties have been selected, the NRI sampling points can be used to locate RSN sample sites. The procedure used for the current RSN cannot be used, however, since most PSU's of the NRI have exactly three sampling points and some have only one sampling point. Thus, it is no longer feasible to center an RSN cropland site about two cropland sampling points. Instead, if a cropland point is selected for the location of a cropland sample site, it is suggested that the site be a square 10-acre site centered at the selected RSN cropland point. If such a site is not all cropland, percent cropland will be noted and specimens taken and kept separately for each stratum.

Efficient sampling within the selected counties could result from careful stratification within the sample counties. The NRI sampling points within a county could first be stratified into cropland points and noncropland points, to insure adequate representation of each of these land types and because agricultural chemical residues are more likely to be found in cropland. Local land use characteristics similar to those suggested for constructing county size measures could be used to further stratify both the cropland points and the noncropland points. Finally, greater selection probabilities would be used in strata where toxic substance residues are more likely to be found. Moreover, at least one cropland site and one noncropland site should be selected from each sample county that contains at least one NRI cropland and one noncropland sample point.

1.3.3 Design Option Three

1.3.3.1 Background

The target population for the National Soil Monitoring Program (NSMP) was the land in the conterminous United States, divided between the Rural Soils Network (RSN) and the Urban Soils Network (USN). Descriptions of these networks are given elsewhere.[11] Both networks were interested in "levels" i.e., the absolute amount of pesticide in the soil, and "trends," the change in this amount with time.

Review of the data indicates large numbers of zero valued observations, and relatively few positive observations. This analytical challenge has been discussed elsewhere [See Lucas et al, Recommendations for the National Surface Water Monitoring Program for Pesticides. Report No. RTI/1864/01-02I]. The conclusion of that analysis was that the appropriate measures of "level" are:

(1) The proportion of positive detections, that is, the relative frequency of last stage sampling units positive for the substance(s) under investigation, and

(2) The proportion of sampling units containing concentrations of substance above some specified level. This level may signal the existence of an undesirable situation.

(3) The geometric mean of the positive values which is a useful concommittant to the data, identifying situations where, for example, the proportion of positive sampling units remains constant, but the level of concentration of toxic substance increases or decreases.

(4) Related to (3), measures based on a truncated, or censored, lognormal model may prove useful.[12]

In the following sections, a two-stage design is proposed, and each stage of sampling is described in some detail. Simple cost and variances are included as means of investing the effect and expense of various alternative sample allocations.

1.3.3.2 Overview of the Proposed Sample Design

The proposed design is a two-stage area probability sample with stratification of the sampling units at each level. The first stage or primary sampling units (PSU's) are counties. The 3141 counties in the United States in aggregate constitute the total land area of the country. Geographic stratification is provided by the four Census Regions. Allocation of PSU's to these regions is in proportion to the land area eligible for the study.

---

[11]National Soils Monitoring Program: Preliminary Report. January, 1980. Research Triangle Institute. EPA Contract No. 68-01-5848.

[12]Owen and DeRouen. Estimation of the mean for lognormal data containing zeros and left-censored values, with application to the measurement of worker exposure to air contaminants. Biometrics: 36:707 (1980).

The question of land area eligibility is currently defined by the membership requirements of the RSN and the USN. It may be advantageous from administrative as well as fiscal and statistical grounds to combine the activities of the soil networks, and consider SMSA counties as a stratum within the survey. This point requiring further review beyond the scope of this study is not addressed. Initial investigation does suggest that savings may reasonably be anticipated. Further discussion is limited to tasks assigned to the RSN.

With the extension of monitoring responsibility from pesticides to toxic substances in general, some revision of the approach seems indicated. The following stratification variables are therefore proposed in addition Census Regions for the PSU's:

(1) Land area,
(2) Population density,
(3) Agricultural activity, and
(4) Industrial activity.

Second stage sampling units (SSU's) are 10-acre plots. These are proposed as the final stage units or analysis units on the assumption that they are sufficiently homogeneous that the effects of subsampling are negligible. This is a verifiable proposition. The problem with SSU's this small is the ability to locate them in the field. The requirement for exactly locating plots is exacerbated by the absence of identifiable boundaries, rendering the task most difficult. To ease this difficulty, Census enumeration districts (ED's) are proposed as readily identifiable segments. The problem is reduced to locating the SSU within the ED, or any suitable subsegment adopted to facilitate matters.

SSU's will be allocated equally to PSU's. A detailed field protocol will locate the points for specimen collection, leaving the minimum of discretion for the field personnel in the selection of these sites. The protocol would specify a grid locating multiple specimen collection sites. The soil collected in a given plot would be composited, unless the homogeneity of the 10-acre plot is under investigation.

Temporal effect is not considered. It is assumed for establishing budget only that one collection per site per year will be made. However, it does not seem reasonable that all toxic substances persist in soils at stable levels throughout the year. This may be satisfactory for heavy metals, particularly at poorly drained sites, but most pesticides dissipate through leaching, transpiration and degradation following application, and volatiles in all likelihood leave the soil almost immediately. Thus, special studies of this phenomenon are recommended over and above the monitoring effort.

1.3.3.2.1 The First Stage Sample

The first stage sampling units are counties, which are often used as sampling units in national surveys. They are easily identified and are political units of sufficient size that a great deal of information is available about them. Indeed, in order to enhance the efficiency of the proposed design, it is recommended that extensive collection of

information be undertaken for each county in the U.S. This information should include:

1) Total land area
2) Cropland and non-cropland acreages, or their estimates
3) Soil maps, characteristics - pH, organic content, etc.
4) Drainage areas and water ways
5) Weather, climatic and meteorologic data
6) Location and size of urban areas
7) Cropping patterns, major crop(s)
8) Location and types of industrial activities, including storage sites
9) Location of dump sites

Moreover, the Master Area Frame maintained by the USDA should be consulted for design information, as well as States with mandatory pesticide reporting laws.

The size measure for the Census Region is its eligible land area. Other measures correlated with toxic substance use do not appear feasible at this level in view of the variability in land use. The present proposal uses the definitions of the RSN to determine eligibility. The number of counties (PSU's) allocated to each Census Region is in proportion to its size, with at least one PSU selected from each region. The allocation of PSU's to further strata is carried on in this fashion with the limitations that there must be at least one PSU in each stratum.

PSU's in each stratum will be selected with probability proportional to size (PPS) and with replacement. As before, the size measure is the land area eligible for the RSN.

It is anticipated that the investment in the collection of the county level information will provide substantial gains in precision through effective stratification. The purpose of this stratification will be to locate regions of approximately equal risk of exposure to toxic substances, hence permit the effective location of sample sites.

Two points can now be made:

(1) The most effective variables for stratification will change for different classes of toxic substances, and may change from substance to substance, and

(2) It is not possible to anticipate which substances will be of major interest in the future.

This leads to the conclusions:

(a) Information may be profitably collected for every county in the United States, and

(b) Any proposed design should be as flexible as possible.

Point (a) supports point (b) above by simplifying the process of making design changes if and when they become necessary. Additionally, the selection of stratification variables which appear to be both general and effective offers the possibility of achieving a flexible and efficient design over the near future.

The approach is to propose the selection of PSU's according to the general stratification scheme which is found most effective at the time of the adoption of the design. These PSU's would then establish the monitoring network (RSN). The selection of the SSU's within the given PSU's according to the procedure below would then determine the specific soil specimen sites. However, it is proposed that the stratification variables within the PSU's, and hence the soil specimen sites, be allowed to change in response to changing interest in toxic substances. It is intended by this technique to maximize the probability of positive results to monitoring efforts.

1.3.3.2.2 The Second Stage Sample

The secondary sampling units (SSU's) will be 10-acre plots. Equal numbers will be selected within each PSU. It is possible that there will be more strata within PSU's than sampling units. This suggests that stratified random sampling will not apply. There are a number of related methods which can be used in this situation.

One procedure is to use a composite index combining several stratification variables. In effect, two or more strata are combined and a 'weight' is assigned to each observation in the new stratum based on the relative sizes of the original strata. Observations are then selected from the new strata by the usual probability methods.

A second procedure is to consider the effect of combinations of the strata and assure at least one observation from important combinations is selected. This can be accomplished by employing the lay-out of an experimental design as if the strata were treatment levels. The latin square is used in this fashion. For example, consider the following case with two stratification variables each at 3 "levels."

Table A. A Latin Square Selection Scheme

		Geographic Location		
		Type 1	Type 2	Type 3
Soil Types	Type 1		x	
	Type 2			x
	Type 3	x		

x = selected plot.

Here with a sample of 3 plots we have observations from each type of location (possibly classified by potential exposure) and of each soil

type. This can be done by: Choosing a "cell" (Soil Type x Location Type Combination) at random, then eliminating the remaining cells in the same row and column from further consideration. A second selection is made at random from the cells in the remaining rows and columns. The row and column containing the second selection are then eliminated and the next random choice is made. This procedure is continued until all the rows and columns are eliminated.

A third procedure generalizes the approach above and is called "controlled selection". The typical use of this procedure is to visual the sample in a tabular array as:

Table B. Example of Controlled Selection

		Geographic Location				
		Site 1	Site 2	Site 3	Site 4	Total
Soil Type	Type 1					$n_{1.}$
	Type 2					$n_{2.}$
	Type 3					$n_{3.}$
	Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	n

Here the total number of plots assigned to a PSU, say, is n. The constraint, or "control", imposed is that the margins of the table, the row and column totals (or proportions if preferred), be satisfied. So, Site 1 must appear $n._1$ times and Soil Type 2 must appear $n._2$ times, and so on. Any arrangement of the sample among the table cells which satisfies these constraints is acceptable. And, at least conceptually, every such arrangement, or a specified subset, is written down and a probability assigned to it. Then one of these arrangements is selected by chance according to the assigned probability.

The complication introduced by this method is the loss of the ability to obtain simply an estimate of precision. The level of control requires either replication to obtain a variance estimate or some approximation be used.

The methodology adopted for the design will depend on the actual stratification variables and the constraints on selection which seem most effective. An important statistical consideration is that the procedure used should provide an unbiased estimate of the PSU parameter of interest (total, mean or proportion). In addition, a measure of precision of the estimate should be capable of reasonable approximation.

1.3.3.3 Size and Allocation of the Sample

Sample size is determined by the level of precision needed to answer the question or questions which are the reason for undertaking a survey. The allocation of the sample is dependent upon locating sources of variation entering the survey and the cost of controlling them. Of

course, these two considerations are interdependent and cannot be solved separately. In order to examine this quantitatively, models approximating cost and variability are constructed. These models are only intended to indicate values depending upon circumstances which may change, but still permitting more rational decision-making rather, than an attempt at an exact description of budget or variability.

1.3.3.3.1 A Cost Model

The total cost of a survey depends upon both fixed and variable costs. Fixed costs are overhead costs which are essentially independent of the sample size - materials, rental of quarters, preparatory work, staff salaries, and so on. Variable costs are unit costs - specimen collection, travel, shipping, etc. For our two stage sample, we assume a simple linear cost model,

$$C = C_0 + C_1 n_1 + C_2 n_1 \bar{n}_2,$$

where

C is the total cost of the survey

C_0 is the fixed cost

C_1 is the variable cost for county-level data

C_2 is the variable cost for plots

n_1 is the number of counties in the sample

$\bar{n}_2$ is the average number of plots per county.

The development of the costs is shown in Table 1.3.3.1. These costs are estimated from related efforts and are only approximate. Different methods in contracting and operating the survey will significantly alter these costs. For example, cooperative agreements with the Department of Agriculture or other interested agencies may produce substantially different field costs. Also laboratory costs are included for "organo-pesticides" and heavy metals. However, different budgeting may appropriately exclude part or all of these costs.

Under the assumptions given we find

C_0 = \$367,800

C_1 = 3,280

C_2 = 926 .

Since the overhead cost includes the collection of preparatory data, maps, etc., on all 3141 counties in the United States, this cost is not included. It may be preferable to:

Table 1.3.3.1 Construction of the Cost Model

1. Selection of counties - first stage units

Item	C_0 - Overhead Costs	C_1 - per County Costs
Construct Frame	$300,000*	
Stratify Frame	1,000	
Develop Size Measures	300	
Select Sample Counties	5,000	
Develop Computerized Data Sets	2,500	
Administration	3,000	100

2. Selection of plots - second stage units

Item	C_0	C_1	C_2 - per Plot Costs
Construct Frame	3500	1000	50
Stratify Frame	4000	150	
Form Segments	500	20	
Select Sample Plots	3000	10	1

3. Field Work and Analysis

Collection of specimens	12000	2000	300
Laboratory Analysis**			560
Data handling	1000		5
Stat. Analysis, Reporting	20000		10
General Administration	10000		

Total: C_0 = $367,800 C_1 = $3,280 C_2 = $926 .

*Includes preparing materials on 3141 counties.

**Uses RTI costs, does not include analysis of toxic substances beyond pesticides and heavy metals.

(1) Do only a subset of the counties, or
(2) Spread this cost over several years.

Ignoring this factor is equivalent to using the cost equation

$$C - C_0 = C_1 n_1 + C_2 n_1 \bar{n}_2$$

which clearly does not affect the relative allocation of the sample. Using the first equation, the estimate cost of a survey of 57 counties with an average of 18.73 plots per county is

$$C = \$367,800 + \$3,280\ (57) + \$926\ (57)\ (18.73)$$

$$= \$1,543,367.$$

1.3.3.3.2 Sample Size Calculations

A minimum acceptable precision must be specified to insure the adequacy of the survey results. The statement "I must know the amount within 10 percent," or "The error in the proportion reported must not exceed 20 percent," specifies a sample size under a particular survey of a proposed study if the heterogeneity of the population under investigation is known.

For the purpose of discussion, the parameter of interest is taken to be the proportion, p, of land (specifically of 10 acre-plots) containing detectable levels of toxic substance. The variance model for the estimator $\hat{p}$ of p is

$$\mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n_1 \bar{n}_2} \{1 + \rho(\bar{n}_2 - 1)\},$$

where ρ is the correlation among plots within a county,

n_1 is the number of counties

$\bar{n}_2$ is the average number of plots per county.

The term in brackets is called the "cluster effect", and it is convenient to write

$$d_c = 1 + \rho(\bar{n}_2 - 1)$$

This model ignores stratification and unequal weighting for simplicity.

The sample allocation problem is choose the number of counties, n_1, and the number of plots, n_2, within counties. For a given budget (which fixes the total sample size), are we wiser to include many counties with few plots per county, or fewer counties with more plots per county? The solution is to balance considerations of cost and variability, that is,

Table 1.3.3.2 Cluster Effect for Selected Values of ρ and $\bar{n}_2$

Pesticide	Intracluster Correlation ρ	Average Number of Plots per County $\bar{n}_2$ 5	10	15	20	25
	0.01	1.04	1.09	1.14	1.19	1.24
	0.06	1.24	1.54	1.84	2.14	2.44
Endrin	0.125	1.50	2.13	2.75	3.38	4.00
Chlordane	0.169	1.68	2.52	3.37	4.21	5.06
Aldrin	0.231	1.92	3.08	4.23	5.39	6.54
Dieldrin	0.298	2.19	3.62	5.17	6.66	8.15
P,P'-DDE	0.430	2.72	4.87	7.02	9.17	11.32

Cluster Effect $d_c = 1 + \rho(\bar{n}_2 - 1)$

Table 1.3.3.3 Minimum Cost Allocation Subject to the Constraint:

$$c.v. = \sqrt{V(\hat{p})/p} \leq 0.10$$

ρ	Average Cluster Size $\bar{n}_2$	Cluster Effect d_c	p = 0.0001 n_1^*	p = 0.0001 Est. Cost**	p = 0.001 n_1	p = 0.001 Est. Cost	p = 0.01 n_1	p = 0.01 Est. Cost	p = 0.10 n_1	p = 0.10 Est. Cost
.01	18.73	1.18	3141	$64,779,986	3141	$64,779,986	622	$12,828,060	57	$1,175,928
.06	7.45	1.39	3141	31,970,880	3141	31,970,880	1843	18,759,020	168	1,710,392
.125	4.98	1.50	3141	24,786,972	3141	24,786,972	3141	24,786,972	271	2,138,980
.169	4.17	1.54	3141	22,431,228	3141	22,431,228	3141	22,431,228	332	2,370,544
.214	3.61	1.56	3141	20,802,394	3141	20,802,394	3141	20,802,394	389	2,576,024
.298	2.89	1.56	3141	18,707,782	3141	18,707,782	3141	18,707,782	487	2,900,242
.430	2.17	1.50	3141	16,614,096	3141	16,614,096	3141	16,614,096	623	3,295,392

The entries in the table were calculated from the formulas:

$$\bar{n}_2 = \left(\frac{C_1}{C_2}\,\frac{1-\rho}{\rho}\right)^{\frac{1}{2}} \qquad d_c = 1 + \rho(\bar{n}_2 - 1) \qquad n_1 = (1 - p)\, d_c \,/\, p\bar{n}_2(c.v.)^2$$

*n_1, the number of counties in the sample, cannot exceed the total number in the United States.

**Estimated Cost does not include the fixed portion, C_0, in the cost equation (see accompanying text)

$$\text{Cost} = C_0 + C_1 n_1 + C_2 n_1 \bar{n}_2$$

and $\bar{n}_2$ = average number of plots per county
C_1 = cost for first stage units = $3280
C_2 = cost per second stage units = $926
p = proportion of land area containing detectable levels of toxic substance.

Table 1.3.3.3 (continued) Minimum Cost Allocation Subject to the Constraint:

$$\text{c.v.} = \sqrt{V(\hat{p})}/p \leq 0.15$$

ρ	Average Cluster Size $\bar{n}_2$	Cluster Effect d_c	p = 0.0001 n_1^*	p = 0.0001 Est. Cost **	p = 0.001 n_1	p = 0.001 Est. Cost	p = 0.01 n_1	p = 0.01 Est. Cost	p = 0.10 n_1	p = 0.10 Est. Cost
.01	18.73	1.18	3141	$64,779,921	2797	$56,685,272	277	$ 5,712,842	25	515,599
.06	7.45	1.39	3141	31,971,296	3141	31.971,296	820	8,346,534	74	753,223
.125	4.98	1.50	3141	24,787,138	3141	24,787,138	1325	10,456,311	120	946,977
.169	4.17	1.54	3141	22,431,200	3141	22,431,200	1624	11,597,666	147	1,049,788
.214	3.67	1.56	3141	20,802,403	3141	20,802,403	1901	12,590,056	172	1,139,181
.298	2.89	1.56	3141	18,708,235	3141	18,708,235	2375	14,145,832	215	1,280,570
.430	2.17	1.50	3141	16,614,068	3141	16,614,068	3041	16,085,126	276	1,459,879

Table 1.3.3.3 (continued) Minimum Cost Allocation Subject to the Constraint:

$$\text{c.v.} = \sqrt{V(\hat{p})}/p \leq .20$$

ρ	Average Cluster Size $\bar{n}_2$	Cluster Effect d_c	p = 0.0001 n_1^*	p = 0.0001 Est. Cost **	p = 0.001 n_1	p = 0.001 Est. Cost	p = 0.01 n_1	p = 0.01 Est. Cost	p = 0.10 n_1	p = 0.10 Est. Cost
.01	18.73	1.18	3141	$64,779,921	1573	$32,441,520	155	$ 3,196,716	14	$ 288,735
.06	7.45	1.39	3141	31,971,296	3141	31,971,296	461	4,692,350	41	417,326
.125	4.98	1.50	3141	24,787,138	3141	24,787,138	745	5,879,152	67	528,729
.169	4.17	1.54	3141	22,431,200	3141	22,431,200	914	6,527,257	83	592,737
.214	3.61	1.56	3141	20,802,403	3141	20,802,403	1067	7,079,837	97	642,417
.298	2.89	1.56	3141	18,708,235	3141	18,705,235	1335	7,951,446	121	720,692
.430	2.17	1.50	3141	16,614,068	3141	16,614,068	1710	9,004,908	155	819,860

the budget goes further if we sample the less expensive units, however precision is improved if more of our observations come from the most variable units (since in the extreme case, if the units all have identically the same value, one observation is sufficient to tell us everything about these units).

Using, the cost and variance equations above we find the values of n_1 and $\bar{n}_2$ which optimize precision for a fixed cost are

$$\bar{n}_2 = \sqrt{\frac{C_1(1-\rho)}{C_2\rho}}$$

and

$$n_1 = \frac{(1-\rho)d_c}{\rho\bar{n}_2 (c.v.)^2}$$

$(c.v.)^2$ is the square of the coefficient of variation or the relative variance. It is the level of precision specified as necessary for this survey, and is given by the equation

$$c.v. = \sqrt{\frac{Var(\hat{p})}{p}}$$

The optimal allocation and the associated cost is given for a range of values of ρ, most of which represent national average values for some of the common pesticides reported in the RSN. These values of ρ are indicated in Table 1.3.3.2 along with the effect of cluster size on d_c, the cluster effect, and the names of the pesticides involved. Table 1.3.3.3 displays the minimum cost allocation and the estimated cost corresponding to these values of ρ, the correlation of the pesticide concentrations within counties. Values of the coefficient of variation (c.v.) on the order of 10 percent are commonly accepted.

1.4 Present RSN Operations

The operational design of the Rural Soils Network (RSN) specified that each site would be randomly designated as a first-year, second-year, third-year, or fourth-year sample site, so that sample specimens would be obtained for one-fourth of the sites in each State during each fiscal year. Specimens were to be obtained at each site no less than once every four years and not more than once per year. Soil specimens were obtained by compositing fifty soil cores, each 2 inches in diameter by 3 inches in depth. The procedure for collecting and compositing these cores and for collecting crop specimens is described in detail in the PPC Division Memorandum 804.3, which is dated April, 1969, and is entitled "Guidelines for Collecting Sample for the National Soil Monitoring Program --1969." This memorandum specifies that soil and crop specimens are to be obtained simultaneously at or shortly before harvest time for the cropland sample. It also specified water and sediment specimens should be collected from the nearest pond to each RSN site, within one mile, four times at equal intervals during each sampling year.

The above operational design appears to have been implemented, except that specimens from ponds have been collected in only one fiscal year, 1973. Moreover, data collection ceased with fiscal year 1975, and very little second round data for assessing trends is available.

1.5 Alternate Operational Design for the RSN

The operational design of the Rural Soils Network (RSN) was well conceived for monitoring agricultural pesticides and herbicides in rural soils, harvested crops, and rural ponds. Some modifications appear, however, to be warranted at this time.

The operational design of the RSN specified that soil and crop specimens be obtained simultaneously at or shortly before harvest time. This data was to be used to monitor levels of compounds in soils and crops, as well as establish relationships between soil and crop residues. Crop specimens should be obtained at or shortly before harvest, since it is the harvested crop that will be consumed. However, harvest time may be less than ideal for obtaining soil specimens. Much pesticide and herbicide residue may often be leached out of or vaporized from the cropland soil by harvest time. This could explain in some measure the preponderance of less than detectable residue levels in the cropland soil data collected thus far (See Section 1.7).

Thus, it may be preferable to obtain cropland soil specimens early in the growing season. It would then be necessary to carefully specify where the soil cores were selected, e.g. on a map of the sample site, so that crop specimens could be obtained near harvest time at practically identical locations.

Noncropland soil specimens could be obtained whenever convenient during the sampling year, since there appears to be no major national relationship between annual seasons and toxic substance residues in noncropland soils. Random points in time are preferable, but may not be logistically feasible. However, the purposive selection a single point in time opens up the opportunity for introducing serious bias. Whatever protocol is adopted, it is important that the protocol be applied uniformly across the nation so that the population being sampled is as well-defined as possible. Sampling some areas when levels of toxic substances are suspected to be high, but not doing so in other areas, would lead to difficulties when making other than local inferences.

Changes in the definition of an RSN sample site that would make its boundaries more readily identifiable would be useful. This would be useful so that the selected sample site could be accurately identified, and the identical site could be revisted periodically to establish trends in residue levels. If the selected site is not precisely defined, the value of the sampling design is lessened. Analyses of trends based upon paired differences may lead to spurious results.

The use of a sample site larger than 10 acres may make it easier to identify site boundaries. However, compositing of the specimens collected at a site is only justifiable if the site is homogeneous with respect to data items. Thus, a fairly small sample site is required if the specimens are to be composited. The alternative would be to report multiple specimens individually.

The use of less than fifty soil cores at a sample site could reduce the expense of collecting specimens and should be considered. The use of a large number of cores is advisable, however, if the cores are to be composited. This insures that the composite is representative of the site by reducing the influence of individual cores. If multiple specimens were to be reported separately within a sample site, fewer cores might be sufficient. An experimental study could be designed to investigate optimal size of sample site and optimal number of soil cores.

Elimination of pond water and pond sediment specimens is probably necessary to keep the cost of the RSN data collection reasonable. The operational design specified that pond specimens were to be obtained four times at equal intervals during each fiscal year for RSN sample sites with a pond within one mile. This procedure is commendable since the pesticide level in pond specimens would probably vary greatly, depending upon the turbidity of the water, the water level and the season. The four equally spaced samples would allow compensation for this variability. Unfortunately, this sampling protocol would probably require a field crew devoted entirely to sampling pond water. Two reasonably spaced collections of pond specimens for each sample site in some sampling years may be worth considering. The pond specimens could be collected early and late in the growing season, possibly simultaneously with the collection of soil specimens and crop specimens, respectively.

Finally, it is important that tests for all toxic substances for which inferences are desired be performed on all sample specimens. This may have been the intention in the past, but the data in Section 1.7 show clearly that some classes of compounds were more regularly tested than others. All compounds for which statistical inferences are desired should be tested in all sample specimens. This requirement may place a practical limit on the number of classes of compounds that can be monitored.

1.6 Recommended Modifications

Since the most cost-effective strategy for modifying the RSN depends to some extent upon information which is not available, the following are simply indications of a way to enhance program efficiency. Design Option 1 seems to have little to recommend it. Its importance lies in its connection with the historical series reflecting the operation of the RSN from FY 1968 to FY 1973. However, given the inactivity of the RSN in the intervening years, there is reason to believe the network would require substantial up-dating which in itself adversely affects the relationship between the RSN and the historical series. Moreover, it may be possible to safeguard the series by appropriately managing the transition to a new network.

Design Option 2 may be the most feasible economically. If a cooperative agreement can be reached with the officials responsible for the operation of the National Resources Inventory (NRI), then the field costs may be kept down. Since the NRI is intended to produce national estimates of various kinds, it is likely to do so for toxic substances in an adequate fashion, and a subsample satisfactory for monitoring purposes.

Design Option 3 represents a monitoring effort geared toward toxic substances specifically. It is expected to perform well in providing the desired data. Should an advantageous cooperative agreement with USDA or others not be obtainable, then this would seem to be the option of choice. And, in fact, it is not impossible that conditions may dictate that a combination of Design Options 2 and 3 be adopted. An economical national estimate may be provided by the NRI network, and may be profitably supplemented by local or special studies based on Design Option 3.

1.7 Statistical Findings and Charts for the RSN

1.7.1 Introduction

Data collection for the Rural Soils Network (RSN) occurred between fiscal year 1968 and fiscal year 1975. The design specified that one-fourth of all sites in each State would be sampled in each year. However, the first year of sampling was regarded as a large scale pilot study and only six States were sampled. The RSN was never fully implemented; the yearly data collection effort is summarized in Table 1.7.1. This table indicates, for example, that the random one-fourth of the cropland sites in Maine that were designated to be first-year cropland sites were sampled in fiscal years 1968 and 1973. It is apparent from Table 1.7.1 that only one-fourth of the noncropland sites have been sampled in most States. Also, most States have a follow-up sample at approximately a four year interval for only one-fourth of the cropland sites. Finally, it is apparent that very little data have been collected for the Mountain Census Division of the United States, possibly because of the expense of collecting data in this region.

In preparation for data analysis, the EPA computer records for the RSN were checked for logical inconsistencies. Twenty-three were found. The methods of identifying and resolving these inconsistencies are discussed in Appendix E. Appendix E also describes the creation of a data set with a structure that more readily lends itself to data analysis than do the EPA data files.

1.7.2 Sampling weights

Proper analysis of the RSN data must account for the characteristics of the sampling design by the use of sampling weights. Sampling weights are adjustments attached to each observation of a data set which usually reflect the probability of selection of the observation. In the case of simple random sampling, the use of weights is quite straightforward. If one individual in a 1000 is randomly selected, i.e., the probability of selection is 1/1000, then each individual "represents" 1000 others and his income, say, is multiplied by 1000 to estimate the total income of 1000 individuals. In more complex survey designs, the same approach applies although the details become more complicated.

The weights for the Rural Soils Network (RSN) depend on two phases of sampling: (1) The selection of the sampling points for the 1967 Conservation Needs Inventory (CNI), and (2) the subsample of the 1967 CNI points selected to locate the RSN sample plots. Therefore, the

Table 1.7.1: Fiscal Years of Data Collection for the Rural Soils Network**

Census Division / State	Cropland Samples						Noncropland Samples			
	Year in Round 1				Round 2		Year in Round 1			
	1	2	3	4	1	2	1	2	3	4
New England										
Maine	68*	69	70	72	73	74	68*	69	70	72*
New Hampshire	69	70	72	73	74		72*			
Vermont	69	70	72	73	74		72*			
Massachusetts	69	70	72	73	74		72*			
Rhode Island	69	70	72	73	74*		72*			
Connecticut	69	70	72	73	74		72*			
Middle Atlantic										
New York	69	70	72	73	74		72*			
New Jersey	69	70	72	73*	74		72*			
Pennsylvania	69	70	72	73	74		72*			
East-North Central										
Ohio	69	70	72	73	74		72*			
Indiana	69	70	72	73	74		72*			
Illinois	69	70	72	73	74		72*			
Michigan	69	70	72	73	74		72*			
Wisconsin	69	70	72	73	74		72*			
Pacific										
Washington	68*	69	72	73	74		68*	69		
Oregon	72	73	74							
California	69	70	72	73	74					
West-North Central										
Minnesota	70									
Iowa	69	70	72	73	74		69			
Missouri	69	70	72	73	74					
N. Dakota	69									
S. Dakota	69	70	72	73	74					
Nebraska	68*	69	70	72	73	74	68*	69		
Kansas	75*									

(continued)

Table 1.7.1: Fiscal Years of Data Collection for the Rural Soils Network (continued)

Census Division / State	Cropland Samples						Noncropland Samples			
	Year in Round 1				Round 2		Year in Round 1			
	1	2	3	4	1	2	1	2	3	4
South Atlantic										
Delaware	69	70	72	73	74		72*			
Maryland	69	70	72	73	74		69	70	72*	
Virginia	68*	69	70	72	73	74	68*	69	70	72*
W. Virginia	69	70	72	73	74		69	72*		
N. Carolina	69	70	72	73	74		72*			
S. Carolina	69	70	72	73	74					
Georgia	68*	69	70	72	73	74	68*	69		
Florida	69	70	72	73	74					
East-South Central										
Kentucky	69	70	72	73	74		72*			
Tennessee	69	70	72	73	74					
Alabama	69	70	72	73	74					
Mississippi	69	70	72	73	74					
West-South Central										
Arkansas	69	70	72	73	74					
Louisana	69	70	72	73	74					
Oklahoma	69	70	72	73	74					
Texas	75*									
Mountain										
Montana	75*									
Idaho	68*	69	72	73	74		68*	69		
Wyoming	69									
Colorado	69									
New Mexico	69									
Arizona	69						69			
Utah	69									
Nevada	69									

* These data are not on the computer files supplied by EPA.

** Source: Personal communications with and computer files supplied by EPA Field Studies Branch, Washington, D.C.

selection probabilities will be discussed which accompany the sampling units in each phase.

1.7.2.1 Sample Selection for the CNI

The CNI is a highly stratified area probability sample, and its sampling weights are rather easily determined. Since stratification requires that units be selected in each stratum (subdivision of the population), there is no choosing among strata. If States are strata, we must draw a sample in every State. If we stratify by county, we sample in every county, and if townships and parts of townships are also strata then we must sample in every such stratum. So there is no selection probability to calculate for strata since each stratum has a 100 percent chance of being selected. Within strata, primary sampling units (PSU's), usually 1 or 2, were selected purely by chance, i.e., at random with equal probabilities.

As discussed in Section 1.2.2, all counties of the conterminous United States that were not entirely urban, were divided into townships and sections, or pseudo-townships and pseudo-sections. The standard sampling procedure used strata composed of 12-section blocks (1/3 of a township), and one quarter-section (the PSU) was drawn at random. Hence, the probability of selection was 1/48, a sampling rate of approximately 2 percent.

Within each PSU, sample "points" were selected by use of a perforated template, which was spun to locate sampling points in an unbiased manner. The perforations formed a grid pattern which was marked on an aerial photograph of the PSU. The CNI sample collected data at each of these sampling points. Among the information collected was land use data, which was used by the RSN to classify each point as either cropland or noncropland. Due to differences in PSU sizes and shapes and the spin of the sampling template, the number of cropland points, the number of noncropland points, and their total change in an unpredictable, or random, manner. These three quantitites are then random variables that can be used in standard statistical procedures. The RSN used these random variables for estimation of proportions of cropland and noncropland acreage in each of the States of the conterminous United States.

If we use the notation

$$U(i,j,k) = \text{the total number of PSU's in stratum k of county j in State i,}$$

then the probability of selecting PSU ℓ when u(i,j,k) PSU's are selected at random from stratum k is

$$p(i,j,k) = \frac{u(i,j,k)}{U(i,j,k)} \cdot$$

It is shown in Appendix D that the selection of sampling points within the PSU can essentially be ignored. The resulting sampling weight for each of the $n(i,j,k,\ell)$ sampling points in PSU ℓ is then

$$W(i,j,k,\ell,m)^{13} = \frac{1}{p(i,j,k)} \quad \text{for } m = 1, 2, \ldots, n(i,j,k,\ell).$$

1.7.2.2 Sample Selection for the RSN

The RSN is based upon a subsample of the CNI sampling points. It is intended to provide valid estimates for cropping regions and some of the larger States, rather than the county level estimates available from the CNI. The RSN is based upon systematic subsamples, one for cropland points and another for noncropland points, selected from the sampling points of the CNI within each State. Each sampling point selected for an RSN sample is used to locate a 10-acre sample plot.

The RSN cropland sample is based upon a systematic subsample of the CNI sampling points that have been classified as cropland points as detailed in Section 1.2.2. This procedure results in a sample in which the PSU's of the CNI occur essentially with probability proportional to size (PPS), where "size" is measured by the proportion of cropland points within the PSU. Thus, PSU's containing a higher proportion of cropland points are more likely to be selected into the RSN cropland sample.

The following notation is useful for expressing the RSN sampling weights:

$v_1(i,j,k,\ell)$ = number of sample cropland points in PSU ℓ.

$v(i,j,k,\ell)$ = total number of sample points in PSU ℓ.

$r_1(i,j,k,\ell)$ = the cropland ratio for PSU ℓ (adjusted as detailed in Appendix D).

$N_1(i) = \Sigma\, r_1(i,j,k,\ell)$ = Sum of the cropland ratio over all units in State i.

$n_1(i)$ = number of RSN cropland sample sites in State i.

The probability that a PSU of the CNI will be selected into the RSN sample is then essentially proportional to

$$r_1(i,j,k,\ell) \cdot \frac{n_1(i)}{N_1(i)} .$$

---

13 Since 640-acre PSU's were sampled at one-fourth the rate of all other sizes of PSU's, the appropriate weight for these sites is $4W(i,j,k,\ell,m)$.

It is well-known[14] that drawing equal sized samples within PSU's selected with probability proportional to size results in a self-weighting sample, i.e. all ultimate sampling units having the same sampling weight. Essentially the same phenomenon occurs with the RSN samples. Most PSU's of the CNI that are selected into the RSN sample receive exactly one RSN sample plot. Thus, under the fairly broad assumptions detailed in Appendix D, the sampling weights for the RSN cropland sample plots are given by

$$W_1(i,j,k,\ell,m_1)^{15} = v(i,j,k,\ell) \cdot \frac{N_1(i)}{n_1(i)} \quad \text{for } m_1 = 1,2,\ldots, n(i,j,k,\ell).$$

Since the total number of points, $v(i,j,k,\ell)$, within a PSU, is essentially constant for most States, the sample is essentially self-weighting for most States.

Details of the derivation of the sampling weights and implementation of approximate sampling weights are found in Appendix D. The approximate sampling weights were calculated and included in the data set constructed for analysis purposes, which is discussed in Appendix E.

1.7.3 Stratification

The two phase sampling design of the RSN necessarily introduces complexities into the data analysis. The first phase sample, the 1967 Conservation Needs Inventory (CNI), was a deeply stratified design. The second phase sample was the systematic selection of ultimate sampling units from the CNI to locate RSN sample sites. Exact variance formulas for estimates based upon the RSN would be very difficult to derive, and would include components of variance from both phases of the design. As is common practice in this situation, approximate variance formulas were used that capture most of the design effects and provide conservative estimates of variance. The major design effects to be accounted for in the RSN design are the stratification effects derived from the CNI sampling design.

The RSN sampling design was described in detail in Section 1.2.2. The dimensions of the stratification in this design are reviewed in Exhibit 1.7.1. The first dimension of stratification in the CNI, and hence the RSN, consists of the 48 States of the conterminous United States. Within some States, large scale geographic strata were defined. For example, the sandhills of Nebraska were treated as a stratum. The irrigated agricultural areas of many States were treated as strata. Desert areas were treated as strata in many States.

The designation of large scale geographic strata within States was usually accompanied by the use of different sizes of PSU's in the CNI

---

[14] See, for example, Kendall, M.G. and Stuart, A. [1968, pg 195]. The Advanced Theory of Statistics, Vol. 3. Hafner, New York.

[15] Or $4W_1(i,j,k,\ell)$ for 640-acre PSU's. (Recall footnote 1).

Exhibit 1.7.1: Dimensions of the RSN Sample Design*

I. Phase One Sample - 1967 Conservation Needs Inventory (CNI).

A. Dimensions of deep stratification.

1. States of the 48 conterminous United States.
2. Large scale geographic strata, etc., sandhills, irrigated areas, etc.
3. Counties that are not entirely urban (crossed with the large scale geographic strata to form smaller sub-county strata).
4. Townships or pseudo-townships within counties or sub-county strata.
5. Strata generally composed of 48 PSU's each within townships or pseudo-townships.

B. Phase One Sample Selection

1. Usually one PSU was selected from each ultimate stratum.
2. A template was used to assign a randomly aligned two-dimensional sample of SSU's within each sampled PSU (the number of SSU's assigned was usually proportional to PSU size).

II. Phase Two Sample - Rural Soils Network (RSN) subsamples

A. Systematic subsamples of the utlimate sampling units, SSU's, from the first phase sample were used to locate the 10-acre RSN sample sites.

*Source: Documents from and personal communications with both the EPA Field Studies Branch at Washington, D.C. and the Statistical Laboratory at Iowa State University.

sample. The irrigated strata were generally very hetergeneous and were of special interest. Thus, 40-acre PSU's were usually used in these strata. It appears that all 40-acre PSU's were assigned to irrigated strata. In addition, the CNI sometimes employed 160-acre PSU's in the irrigated strata. For analysis of the RSN data, a stratum was defined within each State which consisted of all sites in 40-acre PSU's, as well as all sites in 160-acre PSU's which fell within an irrigated stratum of the CNI. Sites within 40 acre PSU's are given by Tables D-4 and D-5 in Appendix D. The sites in 160-acre PSU's used in irrigated strata are shown in Table 1.7.2.

The sandhills stratum in Nebraska was a homogeneous stratum, and 640-acre PSU's were used throughout. Geographically homogeneous strata, such as desert lands, were also defined in the States of New Mexico, South Dakota, Utah, and Wyoming. Apparently, 640-acre PSU's were used exclusively within these strata as well. Moreover, a geographically homogeneous stratum was also defined in Maine. Both 200-acre and 400-acre PSU's were used in this stratum for Maine. Thus, for analysis of the RSN data, a stratum was defined within each State which consisted of all sites in the 200, 400, or 640 acre PSU's. The sites within these oversized PSU's are given by Tables D-4 and D-5 of Appendix D.

All RSN sites of a State that were not classified as being in either of the two large scale geographic strata just defined were considered to be in the "remainder" stratum of that State. For States that contained PSU's of only one size and no irrigated stratum, all sites were considered to be in the "remainder" stratum, which was then identical to the State stratum itself. All States in Table D-2 and D-3 of Appendix D fell into this category, except for Oregon and Idaho (See Table 1.7.2).

1.7.4 Analysis

Several types of analyses are of interest for the RSN data, notably:

(1) Estimation of base levels for residues of toxic substances,
(2) Estimation of changes in mean levels of toxic substance residues from the first round to the second round of data collection, and
(3) Estimation of relationships between soil and crop residue levels.

The reason for analyzing the RSN data in this study was to obtain a measure of precision of residue data based upon the present data collection effort. It was decided that estimation of base levels of residues would be sufficient. In particular, estimation of levels was undertaken for first first round soil data only.

It was found that the data values for most compounds were predominantly zeros. In fact, Tables 1.7.3 and 1.7.4 list numerous compounds for which no detectable levels were found in the cropland and noncropland soils, respectively.

Table 1.7.2: RSN Sites in Counties Having Both Irrigated and Remainder Strata, but only 160-acre PSU's *

State Name (State Code)	County Name (County Code)	Irrigated Stratum† Site Numbers	Remainder Stratum† Site Numbers
Arizona (04)	Apache (001)		10-13
	Cochise (003)	1	14,15
New Mexico(35)	Curry (009)	3	2
	Hidalgo (023)	5	
	Roosevelt (041)	8	9
	Torrance (057)	10	
Oregon (41)	Crook (013)	78,150	4
	Grant (023)	81,154	8
	Lane (037)	16,17,90,91,162, 163	
	Malheur (045)	20-22,94,96,166,167,169	95,168
Idaho (16)	Ada (001)	1,64	97
	Adams (003)'	127	
	Bannocke (005)		2,65,190
	Bear Lake (007)		3,98,128,191
	Bingham (011)	4,5,193	67,68,99,130
	Blaine (013)		100,131
	Booneville (019)	69,102,133,195,196	7,8,70,132
	Butte (023)	134	103
	Caribou (029)	199	11,12,74,104,137,200
	Cassia (031)	13,75,76,105,202	138,139,201
	Clark (033)	77	14
	Custer (037)		107-109,203
	Elmore (039)		15,78,110
	Franklin (041)		16,141,204
	Fremont (043)	79,80	17,111,142,205
	Gem (045)	143	
	Kootenai (055)	147	84
	Lemhi (059)	211	117,118
	Lincoln (063)	24	
	Madison (065)	213	25,87,150
	Oneida (071)	28,153	90,216
	Owyhee (073)	120,154	91,121,122
	Payette (075)	217	
	Power (077)	93,156	29,30,92,155,218
	Teton (081)	31	94,157
	Twin Falls (083)	32,95,158,220,221	33
	Valley (085)	96	125,126
	Washington (087)	159	222

†Only sites that were surveyed by the RSN have been classified. Classification of all sites in these counties would require considerably more effort.

*Source: CNI site numbers corresponding to the RSN site numbers were obtained from the EPA Field Studies Branch, Washington, D.C. The stratum classification for each of these CNI sites was obtained from the Statistical Laboratory of Iowa State University.

Table 1.7.3: Compounds with No Detectable Levels in Cropland Soils*

Compound	Sample Size
Alachlor	6071
Photodieldrin	6071
Benzene Heptachloride	6071
Mirex	6071
Prolan	2846
Bulan	2846
Gamma Chlordane	37
Folex	2341

*Source: Computer files supplied by EPA Field Studies Branch, Washington, D.C.

Table 1.7.4: Compounds with No Detectable Levels in Noncropland Soils*

Compound	Sample Size	Compound	Sample Size**
Alachlor	238	Bulan	2
DCPA	238	Gamma Chlordane	0
o,p-'TDE	238	Carbophenothion	2
Photodieldrin	238	DEF	2
Endosulfan I	238	Diazinon	2
Endosulfan II	238	Ethion	2
Endrin	238	Folex	2
Endrin Aldehyde	238	Malathion	2
Endrin Ketone	238	Methyl Parathion	2
Heptachlor	238	Ethyl Parathion	2
Isodrin	238	Phorate	2
Lindane	238	2,4-D	1
Benzene Heptachloride	238	Atrazine	9
Methoxychlor	238		
PCNB	238		
Propachlor	238		
Ronnel	238		
Trifluralin	238		
Mirex	238		
Ovex	238		
PCB	238		

*Source: Computer files supplied by EPA Field Studies Branch, Washington, D.C.

**Rarely tested class of chemicals.

It is also evident from these and subsequent tables in this section that some classes of compounds were tested for more regularly than others, which raises questions about what generalizations can be made from this data. It would be of interest to know what criteria were used to determine whether or not a test would be performed.

Moreover the exclusion of some States from the sample restricts the population to which inferences are valid. It can be seen from Table 1.7.1 that nearly complete data exists for some Census Divisions, while there is very little data for others.

The predominance of zero values in the residue data results in J-shaped distributions for the amount of residue detected for most compounds. This type of data presents some analysis problems. For example, the weighted mean of the raw data values has little meaning if most values are zero and a few are large. Thus, some type of data transformation is generally required in order to obtain a meaningful analysis [See Lucas, et al, Recommendations for the National Surface Water Monitoring Program for Pesticides. Report No. RTI/1864/01-02I]. Ideally, each compound should be considered individually to determine an appropriate transformation, if any. A ubiquitous compound like arsenic may not require a transformation. The analysis of the first round soil data was computed on three scales: (1) The raw data, (2) a logarithmic scale, and (3) a proportion scale. The raw data values exceeding the minimum detectable level (MDL) were also analyzed as a separate data set. The results are shown in Tables 1.7.5 through 1.7.9.

Extensive analyses were not considered appropriate for compounds for which there were few detections - observations in excess of the minimum detectable level (MDL). The analyses for these compounds are presented in Tables 1.7.5 and 1.7.6 for cropland and noncropland soils, respectively. Each of these tables contains the following information for the compounds represented:

(1) The sample size, i.e., number of sites for which the presence of the compound was tested,
(2) The number of data values exceeding the minimum detectable level,
(3) The largest amount of the compound detected at any one site in parts per million (ppm), and
(4) The weighted average,

$$\bar{x}_+ = \frac{\Sigma w_i x_i}{\Sigma w_i},$$

of the detections in ppm where the sampling weights are represented by w_i and the detections (amounts exceeding the MDL) are denoted by x_i.

For the analyses on the logarithmic scale, the data values, say x, were transformed to $\log_e (x+1)$. This is a transformation often found to be useful for stabilizing the variances of data that consist of positive

integers covering a wide range.[16] The presence of many zero values for most of the compounds makes this transformation of questionable value for such compounds. For presentation of the findings on this scale in Tables 1.7.7 through 1.7.9, the results have been transformed again to the original scale. In particular, if $\bar{y}$ represents the weighted mean of the log-transformed data, the value reported is given by

$$\bar{x}_g = \text{Antilog}_e (\bar{y}) - 1 ,$$

which bears a strong analogy to the geometric mean. Actually, the geometric mean is identically zero when any of the data values are zero.

For analyses on the proportion scale, all data values above the minimum detectable level (MDL) were replaced by the value one (so that their sum is the number of positive values). The weighted mean on this scale is a weighted estimate of the proportion of the sampled land area with a residue level in excess of the MDL. Since this scale is felt to be the most appropriate for analysis of the residue data, the standard error and the design effect for the estimated proportion are also presented in Tables 1.7.7 through 1.7.9. The statistical approach used for computation of the standard errors and design effects was a first-order Taylor series approximation as implemented in computer software developed by RTI for analysis of nested probability samples [See SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data. Report No. RTI/1789/00-01F].

Estimation of standard errors and design effects required that some of the strata defined in Section 1.7.3 be combined. In particular, strata that received only one sampling unit had to be combined with other strata to produce valid estimates of sampling variances. In order to determine where this was necessary, the RSN records were first sorted by States, by large scale geographic strata within States, and finally by counties within large scale geographic strata (See Exhibit 1.7.1). When a stratum defined by these three levels of sorting (i.e., an individual county portion of a large scale geographic stratum) contained only a single round one soil record, this stratum was placed into a "residual county" stratum created within the large scale geographic stratum. Recall that the States having no large scale geographic stratification can be thought of as a single large scale geographic stratum. Finally, whenever a "residual county" stratum within a large scale geographic stratum consisted of only a single Round One soil record, the stratum identification of the record in this "residual county" stratum was changed to that of an arbitrary county within the same large scale geographic stratum. The goal of this strategy was to achieve the maximum possible benefits from the CNI stratification for estimation of standard errors and design effects.

Since it was not possible to account for all dimensions of the CNI stratification (See Exhibit 1.7.1), the standard errors computed are

---

[16] See page 157 of Steel, R.G.D. and Torrie, J.H. [1960]. Principles and Procedures of Statistics. McGraw-Hill, New York.

Table 1.7.5: Statistics for Compounds with Few Detectable Levels in Cropland Soils for Round One*

Compound	n [1]	n_+ [2]	Max [3]	$\bar{x}_+$ [4]
DCPA	6071	3	1190	632.92
Dicofol	6071	16	2150	370.40
Endosulfan I	6071	7	240	95.83
Endosulfan II	6071	15	1240	172.10
Endosulfan Sulfate	6071	18	2070	343.85
Endrin Aldehyde	6071	1	30	30.00
Endrin Ketone	6071	10	380	98.19
Lindane	6071	21	350	51.92
Methoxychlor	6071	1	280	280.00
PCNB	6071	4	2610	1103.87
Propachlor	6071	5	100	80.27
Ronnel	6071	1	190	190.00
Ovex	6071	1	1130	1130.00
PCB	6071	2	1490	1130.98
Carbophenothion	2341	1	230	230.00
DEF	2341	9	670	272.63
Diazinon	2341	9	170	82.01
Ethion	2341	3	240	107.95
Malathion	2341	5	360	163.26
Methyl Parathion	2341	1	10	10.00
Ethyl Parathion	2341	18	3010	296.05
Phorate	2341	10	400	76.16
2,4-D	188	3	30	17.26

[1] Sample size.

[2] Number of occurrences above the MDL.

[3] Maximum amount detected (PPM).

[4] Weighted average of the data values in excess of the MDL (PPM).

*Source: Computer files supplied by EPA Field Studies Branch, Washington, D.C.

Table 1.7.6: Statistics for Compounds with Few Detectable Levels in Noncropland Soils for Round One*

Compound	$n^{1/}$	$n_+^{2/}$	$Max^{3/}$	$\bar{x}_+^{4/}$
Aldrin	238	1	20	20.00
Chlordane	238	5	500	200.34
o,p'-DDE	238	2	30	24.57
o,p'-DDT	238	8	50	20.43
o,p'-TDE	238	7	180	45.47
Dicofol	238	2	290	138.00
Dieldrin	238	10	90	29.00
Endosulfan Sulfate	238	1	80	80.00
Heptachlor Epoxide	238	2	10	10.00
Toxaphene	238	1	520	520.00

1/ Sample size.

2/ Number of occurances above the MDL.

3/ Maximum amount detected (ppm).

4/ Weighted average of the data values in excess of the MDL (ppm).

*Source: Computer files supplied by EPA Field Studies Branch, Washington, D.C.

Table 1.7.7: Statistics for Compounds with Detectable Levels in Noncropland Soils for Round One*

Compound	Atrazine							
	n[1/]	Max[2/]	$\bar{x}_+$ [3/]	$\bar{x}$ [4/]	$\bar{x}_g$ [5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
p,p'-DDE	238	310	37.02	3.51	0.35	0.09	0.02	0.92
p,p'-DDT	238	230	54.12	3.49	0.26	0.06	0.02	1.11
Arsenic	233	54,170	3,957.92	3,772.27	1,618.71	0.95	0.02	1.32

(continued)

[1/] Sample size.

[2/] Maximum amount detected (ppm).

[3/] Weighted average of the data values in excess of the MDL (ppm).

[4/] Weighted average of the amount detected (ppm).

[5/] Antilog$_e$ (weighted average of log$_e$ (amount +1)-1); analogous to the geometric mean (ppm).

[6/] Weighted proportion of cases with data values in excess of the MDL.

[7/] Standard deviation of the estimated proportion.

[8/] Design effect for the estimated proportion.

*Source: Computer files supplied by the EPA Field Studies Branch, Washington, D.C.

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One*

Census Division	Aldrin							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	13,280	219.65	23.06	.54	0.11	0.00	0.79
New England	72	280	280.00	4.02	.08	0.01	0.01	1.05
Middle Atlantic	296	150	90.89	.60	.03	0.01	0.00	1.01
East-North Central	1595	13,280	277.06	61.89	1.59	0.22	0.01	0.79
Pacific	505	170	54.91	.99	.07	0.02	0.01	1.03
West-North Central	1943	4,250	166.47	17.56	.54	0.11	0.01	0.79
South Atlantic	482	570	123.23	4.10	.14	0.03	0.01	0.89
East-South Central	429	420	110.00	2.76	.10	0.03	0.01	1.10
West-South Central	546	60	20.72	.68	.10	0.03	0.01	0.96
Mountain	203	20	20.00	.10	.01	0.00	0.00	0.99

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Chlordane							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	13,340	645.24	56.74	.63	0.09	0.00	0.93
New England	72	2,200	693.19	43.30	.45	0.06	0.03	1.07
Middle Atlantic	296	3,190	596.78	26.26	.28	0.04	0.01	0.93
East-North Central	1595	6,980	809.89	120.76	1.48	0.15	0.01	0.87
Pacific	505	2,460	527.37	14.71	.15	0.03	0.01	0.97
West-North Central	1943	8,040	489.02	40.96	.55	0.08	0.01	0.86
South Atlantic	482	13,340	655.08	65.30	.68	0.10	0.01	0.89
East-South Central	429	7,890	753.98	35.67	.30	0.05	0.01	0.93
West-South Central	546	260	116.26	1.26	.05	0.01	0.00	1.19
Mountain	203	480	164.88	11.32	.38	0.07	0.03	2.32

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	o,p ' - DDE							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	510	45.90	.98	.07	0.02	0.00	0.83
New England	72	30	30.00	.12	.01	0.00	0.00	0.29
Middle Atlantic	296	100	40.86	1.35	.12	0.03	0.01	1.11
East-North Central	1595	510	109.60	.62	.02	0.01	0.00	1.00
Pacific	505	380	51.53	5.02	.40	0.10	0.01	0.86
West-North Central	1943	90	27.37	.06	.01	0.00	0.00	0.86
South Atlantic	482	140	29.23	2.00	.23	0.07	0.01	0.98
East-South Central	429	80	32.41	1.66	.19	0.05	0.01	0.93
West-South Central	546	250	67.24	1.42	.08	0.02	0.01	1.02
Mountain	203	70	35.00	0.16	.02	0.00	0.00	0.12

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	p,p ' - DDE							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	54,980	303.39	59.68	1.34	0.20	0.00	0.56
New England	72	4,340	343.43	117.12	3.95	0.34	0.05	0.72
Middle Atlantic	296	54,980	1207.38	340.07	2.42	0.28	0.03	0.99
East-North Central	1595	7,160	338.10	27.14	.36	0.08	0.01	0.90
Pacific	505	16,690	374.57	171.82	7.08	0.46	0.02	0.57
West-North Central	1943	550	54.46	3.29	.23	0.06	0.00	0.84
South Atlantic	482	5,410	221.13	129.96	12.87	0.59	0.02	0.79
East-South Central	429	1,710	209.50	106.09	8.87	0.51	0.02	0.55
West-South Central	546	6,210	364.48	97.57	2.46	0.27	0.02	0.64
Mountain	203	840	89.98	17.36	1.10	0.19	0.02	0.66

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	p,p' - DDT							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	245,180	1044.70	187.17	1.51	0.18	0.00	0.56
New England	72	4,650	850.87	253.33	4.81	0.30	0.04	0.56
Middle Atlantic	296	245,180	5890.52	1527.29	2.74	0.26	0.03	0.99
East-North Central	1595	35,920	1610.18	97.31	.34	0.06	0.01	0.93
Pacific	505	19,750	783.42	297.64	6.24	0.38	0.02	0.62
West-North Central	1943	1,420	127.08	7.80	.30	0.06	0.00	0.83
South Atlantic	482	20,260	582.11	318.45	17.36	0.55	0.02	0.73
East-South Central	429	16,070	967.43	478.18	14.64	0.49	0.02	0.52
West-South Central	546	15,860	1002.76	252.82	2.90	0.25	0.01	0.56
Mountain	203	3,230	226.55	38.88	1.05	0.17	0.02	0.76

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	o,p ' - DDT							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	32,750	307.91	35.94	.67	0.12	0.00	0.58
New England	72	860	169.43	38.13	1.80	0.23	0.03	0.45
Middle Atlantic	296	32,750	1552.14	262.43	1.14	0.17	0.02	0.93
East-North Central	1595	8,210	797.32	20.47	.14	0.03	0.00	0.87
Pacific	505	4,510	205.01	56.99	2.35	0.28	0.02	0.74
West-North Central	1943	410	46.63	1.31	.09	0.03	0.00	1.00
South Atlantic	482	4,180	171.33	67.89	4.64	0.40	0.02	0.76
East-South Central	429	1,790	233.35	83.90	4.31	0.36	0.02	0.55
West-South Central	546	5,620	374.76	63.94	1.23	0.17	0.01	0.49
Mountain	203	290	66.11	5.84	.38	0.09	0.02	0.69

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	p,p ' - TDE							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	38,460	349.24	31.78	.46	0.09	0.00	0.73
New England	72	8,200	616.34	156.40	2.35	0.25	0.04	0.70
Middle Atlantic	296	38,460	1978.77	255.99	.82	0.13	0.02	1.00
East-North Central	1595	31,430	859.24	25.67	.14	0.03	0.00	0.95
Pacific	505	20,130	357.52	68.26	1.27	0.19	0.02	0.79
West-North Central	1943	500	32.42	.63	.06	0.02	0.00	1.09
South Atlantic	482	7,470	177.33	63.33	3.57	0.36	0.02	0.85
East-South Central	429	1,250	135.30	31.50	1.64	0.23	0.02	0.94
West-South Central	546	1,670	159.58	21.10	.72	0.13	0.01	0.68
Mountain	203	150	38.19	1.86	.17	0.05	0.01	0.77

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	o,p ' - TDE							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	16,790	387.39	5.71	.06	0.01	0.00	0.86
New England	72	50	50.00	.72	.06	0.01	0.01	1.06
Middle Atlantic	296	16,790	2156.69	91.14	.27	0.04	0.01	0.87
East-North Central	1595	1,300	206.48	1.01	.02	0.00	0.00	0.98
Pacific	505	4,520	252.80	13.76	.27	0.05	0.01	0.93
West-North Central	1943	100	100.00	.04	0.00	0.00	0.00	0.81
South Atlantic	482	1,350	124.03	9.00	.36	0.07	0.01	1.03
East-South Central	429	490	138.89	2.56	.08	0.02	0.01	1.00
West-South Central	546	210	150.00	.20	.01	0.00	0.00	0.37
Mountain	203	10	10.00	.14	.03	0.01	0.00	0.32

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Dieldrin							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	9,830	150.35	41.14	2.22	0.27	0.01	0.84
New England	72	4,640	1,087.94	123.64	.79	0.11	0.04	1.06
Middle Atlantic	296	9,830	284.49	60.22	1.29	0.21	0.02	0.92
East-North Central	1595	6,180	196.21	72.36	4.58	0.37	0.01	0.71
Pacific	505	2,150	126.37	20.69	.93	0.16	0.02	0.92
West-North Central	1943	1,620	113.45	32.92	2.35	0.29	0.01	0.84
South Atlantic	482	1,850	175.57	43.34	1.77	0.25	0.02	0.83
East-South Central	429	650	61.72	13.05	1.08	0.21	0.02	1.04
West-South Central	546	270	70.73	9.42	.68	0.13	0.01	0.66
Mountain	203	610	61.70	11.69	.95	0.19	0.03	1.55

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Endrin							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	2,130	142.72	1.73	.05	0.01	0.00	0.81
New England	72	150	150.00	2.17	.07	0.01	0.01	1.06
Middle Atlantic	296	560	313.43	3.32	.06	0.01	0.00	0.21
East-North Central	1595	20	14.89	.02	0.00	0.00	0.00	0.98
Pacific	505	160	49.22	1.54	.12	0.03	0.01	0.86
West-North Central	1943	80	26.53	.15	.02	0.01	0.00	0.79
South Atlantic	482	2,130	347.11	12.28	.17	0.04	0.01	1.00
East-South Central	429	640	141.47	4.07	.13	0.03	0.01	0.73
West-South Central	546	480	101.57	2.21	.09	0.02	0.01	0.92
Mountain	203	220	33.43	.57	.05	0.02	0.01	0.90

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Heptachlor							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	1,710	101.01	4.78	.20	0.05	0.00	0.81
New England	72	40	25.00	.72	.09	0.03	0.02	1.07
Middle Atlantic	296	10	10.00	.04	.01	0.00	0.00	1.17
East-North Central	1595	1370	102.76	12.23	.57	0.12	0.01	0.84
Pacific	505	20	20.00	.04	.01	0.00	0.00	0.98
West-North Central	1943	1,710	109.97	3.99	.14	0.04	0.00	0.74
South Atlantic	482	340	93.18	1.56	.06	0.02	0.01	1.02
East-South Central	429	70	18.30	.34	.05	0.02	0.01	1.03
West-South Central	546	10	10.00	.02	.01	0.00	0.00	1.30
Mountain	203	260	140.00	.34	.01	0.00	0.00	0.21

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Heptachlor Epoxide							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	1,080	54.59	4.24	.31	0.08	0.00	0.92
New England	72	60	32.64	1.91	.22	0.06	0.03	1.12
Middle Atlantic	296	60	24.75	.83	.11	0.03	0.01	0.86
East-North Central	1595	1,080	69.56	9.21	.65	0.13	0.01	0.84
Pacific	505	70	18.46	.51	.08	0.03	0.01	1.00
West-North Central	1943	330	43.16	3.55	.31	0.08	0.01	0.85
South Atlantic	482	180	41.02	2.97	.27	0.07	0.01	0.94
East-South Central	429	720	96.30	2.70	.11	0.03	0.01	1.03
West-South Central	546	10	10.00	.02	.01	0.00	0.00	1.30
Mountain	203	50	37.65	1.61	.16	0.04	0.02	2.42

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Isodrin							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	180	21.68	.16	.02	0.01	0.00	0.96
New England	72	0	0.0	0.0	0.0	0.0	0.0	1.00
Middle Atlantic	296	0	0.0	0.0	0.0	0.0	0.0	1.00
East-North Central	1595	180	23.17	.51	.06	0.02	0.00	0.99
Pacific	505	0	0.0	0.0	0.0	0.0	0.0	1.00
West-North Central	1943	50	18.99	.08	.01	0.00	0.00	0.82
South Atlantic	482	0	0	0	0	0.0	0.0	1.00
East-South Central	429	10	10.00	.05	.01	0.01	0.00	1.08
West-South Central	546	0	0.0	0.0	0.0	0.0	0.0	1.00
Mountain	203	0	0.0	0.0	0.0	0.0	0.0	1.00

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Toxaphene							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	36,330	3,562.56	129.98	.32	0.04	0.00	0.64
New England	72	0	0.0	0.0	0.0	0.0	0.0	1.00
Middle Atlantic	296	0	0.0	0.0	0.0	0.0	0.0	1.00
East-North Central	1595	0	0.0	0.0	0.0	0.0	0.0	1.00
Pacific	505	8,300	2,225.71	208.16	.99	0.09	0.01	0.76
West-North Central	1943	5,970	3,031.10	5.08	.01	0.00	0.00	0.81
South Atlantic	482	18,100	3,012.79	423.65	1.89	0.14	0.01	0.84
East-South Central	429	21,000	3,460.30	629.80	2.97	0.18	0.02	0.71
West-South Central	546	36,330	7,271.25	519.17	.80	0.07	0.01	0.66
Mountain	203	4,960	3,398.33	47.19	.12	0.01	0.01	0.51

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Trifluralin							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	1,860	99.33	3.20	.14	0.03	0.00	0.80
New England	72	0	0.0	0.0	0.0	0.0	0.0	1.00
Middle Atlantic	296	140	92.95	1.12	.05	0.01	0.01	0.93
East-North Central	1595	600	90.40	2.11	.11	0.02	0.00	0.99
Pacific	505	1,290	159.72	4.05	.11	0.03	0.01	0.97
West-North Central	1943	680	94.74	2.42	.12	0.03	0.00	0.69
South Atlantic	482	1,860	122.55	6.67	.23	0.05	0.01	0.85
East-South Central	429	270	76.00	7.45	.48	0.10	0.01	0.73
West-South Central	546	370	118.86	4.57	.19	0.04	0.01	0.81
Mountain	203	240	97.07	1.90	.08	0.02	0.01	0.91

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Arsenic							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	4690	180,420	5,869.29	5,665.15	2,863.07	0.97	0.00	1.44
New England	59	69,100	10,649.32	10,462.80	4,913.77	.98	0.02	1.05
Middle Atlantic	222	180,420	9,211.71	9,034.18	5,270.13	0.98	0.01	1.11
East-North Central	1191	99,400	6,618.49	6,448.51	3,427.92	0.97	0.00	1.01
Pacific	311	61,810	4,490.05	4,404,59	2,642.87	0.98	0.01	0.96
West-North Central	1598	107,450	5,948.02	5,667.13	2,778.43	0.95	0.01	1.62
South Atlantic	402	25,600	3,251.96	3,080.14	1,260.43	0.95	0.01	0.96
East-South Central	326	34,480	7,286.42	7,180.89	4,768.52	0.99	0.01	0.94
West-South Central	410	33,500	4,138.06	4,072.43	2,391.27	0.98	0.01	1.31
Mountain	171	15,820	3,555.91	3,430.49	1.957.63	0.96	0.01	0.84

(continued)

Table 1.7.8: Statistics for Compounds with Detectable Levels in Cropland Soils by Census Division for Round One* (continued)

Census Division	Atrazine							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	523	16,730	231.40	115.34	8.30	0.50	0.02	1.16
New England	0	-	-	-	-	-	-	-
Middle Atlantic	0	-	-	-	-	-	-	-
East-North Central	235	1,380	137.22	70.21	8.12	0.51	0.03	0.99
Pacific	0	-	-	-	-	-	-	-
West-North Central	288	16,730	303.75	148.45	8.40	0.49	0.03	1.27
South Atlantic	0	-	-	-	-	-	-	-
East-South Central	0	-	-	-	-	-	-	-
West-South Central	0	-	-	-	-	-	-	-
Mountain	0	-	-	-	-	-	-	-

(continued)

[1] Sample size.

[2] Maximum amount detected (ppm).

[3] Weighted average of the data values in excess of the MDL (ppm).

[4] Weighted average of the amount detected (ppm).

[5] Antilog$_e$ (weighted average of log$_e$ (amount +1)-1); analogous to the geometric mean (ppm).

[6] Weighted proportion of cases with data values in excess of the MDL.

[7] Standard deviation of the estimated proportion.

[8] Design effect for the estimated proportion.

* Source: Computer files supplied by the EPA Field Studies Branch, Washington, D.C.

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One

Cropping Region	Aldrin							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	13,280	219.65	23.06	.54	0.11	0.00	0.79
Corn	1386	4,250	192.83	42.83	1.53	0.22	0.01	0.89
Wheat & Small Grains	1056	220	55.25	0.63	.04	0.01	0.00	0.77
Cotton	221	0	0.0	0.0	0.0	0.0	0.0	1.00
Soybeans	1271	13,280	290.73	61.75	1.39	0.21	0.01	0.96
General Farming	699	1,220	167.63	10.75	.31	0.06	0.01	0.87
Hay	609	280	78.28	1.38	.07	0.02	0.00	0.76
Vegetables	557	350	80.69	2.23	.11	0.03	0.01	1.02
Fruit or Nut Orchard	253	470	172.23	3.42	.09	0.02	0.01	0.59

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Chlordane							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	13,340	645.24	56.74	.63	0.09	0.00	0.93
Corn	1386	8,040	652.22	113.85	1.69	0.17	0.01	0.94
Wheat & Small Grains	1056	660	206.17	1.47	.04	0.01	0.00	0.81
Cotton	221	620	264.00	6.30	.14	0.02	0.01	1.06
Soybeans	1271	5,620	736.21	97.13	1.18	0.13	0.01	0.98
General Farming	699	1,190	321.79	15.29	.27	0.05	0.01	0.97
Hay	609	7,890	620.00	25.55	.23	0.04	0.01	1.59
Vegetables	557	13,340	764.19	61.76	.55	0.08	0.01	1.26
Fruit or Nut Orchard	253	2,720	474.67	51.77	.77	0.11	0.02	0.86

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	o,p' - DDE							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	510	45.90	.98	.07	0.02	0.00	0.83
Corn	1386	90	28.65	.26	.03	0.01	0.00	0.99
Wheat & Small Grains	1056	380	121.31	.35	.01	0.00	0.00	0.70
Cotton	221	250	41.70	6.27	.70	0.15	0.02	1.01
Soybeans	1271	200	31.97	.52	.05	0.02	0.00	0.87
General Farming	699	10	10.00	.02	.01	0.00	0.00	0.90
Hay	609	30	20.00	.09	.01	0.00	0.00	0.94
Vegetables	557	140	39.37	1.77	.16	0.05	0.01	0.95
Fruit or Nut Orchard	253	510	63.95	9.57	.70	0.15	0.02	0.97

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	p,p' - DDE							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	54,980	303.39	59.68	1.34	0.20	0.00	0.56
Corn	1386	550	68.21	7.21	.48	0.11	0.01	0.93
Wheat & Small Grains	1056	2,270	127.47	9.62	.32	0.08	0.01	0.61
Cotton	221	6,210	344.08	272.61	52.52	0.79	0.03	0.87
Soybeans	1271	4,760	226.36	53.45	1.83	0.24	0.01	0.56
General Farming	699	4,550	154.69	13.84	.39	0.09	0.01	0.95
Hay	609	8,090	272.45	28.48	.49	0.10	0.01	0.80
Vegetables	557	6,820	222.87	107.16	6.92	0.48	0.02	0.72
Fruit or Nut Orchard	253	54,980	974.93	611.57	21.20	0.63	0.03	0.88

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	p,p' - DDT							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	245,180	1,044.70	187.17	1.51	0.18	0.00	0.56
Corn	1386	3,080	179.52	19.55	.60	0.11	0.01	0.90
Wheat & Small Grains	1056	5,160	218.00	13.67	.31	0.06	0.01	0.66
Cotton	221	15,860	1,144.63	890.55	98.48	0.78	0.03	0.81
Soybeans	1271	16,070	793.83	174.66	2.25	0.22	0.01	0.53
General Farming	699	23,700	707.81	45.21	.32	0.06	0.01	0.93
Hay	609	38,550	847.73	74.62	.48	0.09	0.01	0.75
Vegetables	557	69,300	1,048.12	440.77	8.21	0.42	0.02	0.73
Fruit or Nut Orchard	253	245,180	3,131.82	1,753.51	20.54	0.56	0.03	0.89

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	o,p' - DDT							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	32,750	307.91	35.94	.67	0.12	0.00	0.58
Corn	1386	470	71.41	3.12	.17	0.04	0.01	1.06
Wheat & Small Grains	1056	620	69.77	2.76	.15	0.04	0.00	0.67
Cotton	221	5,620	328.24	203.82	20.33	0.62	0.03	0.80
Soybeans	1271	3,320	212.36	32.55	.97	0.15	0.01	0.46
General Farming	699	3,790	225.16	7.62	.14	0.03	0.01	0.99
Hay	609	14,050	519.05	24.09	.21	0.05	0.01	0.77
Vegetables	557	11,700	279.46	82.86	2.67	0.30	0.02	0.84
Fruit or Nut Orchard	253	32,750	738.55	292.74	5.62	0.40	0.03	0.89

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	p,p' - TDE							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	38,460	349.24	31.78	.46	0.09	0.00	0.73
Corn	1386	1,230	88.79	4.48	.20	0.05	0.01	1.02
Wheat & Small Grains	1056	370	62.22	1.53	.09	0.02	0.00	0.64
Cotton	221	1,670	172.24	75.23	5.82	0.44	0.03	1.00
Soybeans	1271	1,250	123.35	13.46	.55	0.11	0.01	0.75
General Farming	699	2,070	195.38	4.15	.08	0.02	0.01	0.94
Hay	609	8,200	368.31	15.09	.17	0.04	0.01	0.81
Vegetables	557	31,430	494.21	125.08	1.91	0.25	0.02	0.86
Fruit or Nut Orchard	253	38,460	1,155.54	329.80	2.86	0.29	0.03	1.03

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	o,p' - TDE							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	16,790	387.39	5.71	.06	0.01	0.00	0.86
Corn	1386	340	112.20	.70	.03	0.01	0.00	0.97
Wheat & Small Grains	1056	150	46.58	.23	.02	0.01	0.00	0.46
Cotton	221	490	161.17	7.10	.22	0.04	0.01	1.11
Soybeans	1271	210	49.52	.41	.03	0.01	0.00	1.08
General Farming	699	100	67.24	.24	.01	0.00	0.00	0.86
Hay	609	230	100.38	.69	.03	0.00	0.01	0.87
Vegetables	557	4,870	237.14	13.88	.28	0.06	0.01	0.90
Fruit or Nut Orchard	253	16,790	1,265.99	104.07	.52	0.08	0.02	1.04

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Dieldrin							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	9,830	150.35	41.14	2.22	0.27	0.01	0.84
Corn	1386	1,620	149.79	74.68	8.30	0.50	0.01	0.90
Wheat & Small Grains	1056	610	51.14	4.43	.34	0.09	0.01	1.20
Cotton	221	1,280	86.50	12.04	.60	0.14	0.02	1.09
Soybeans	1271	6,180	165.48	64.05	4.58	0.39	0.01	0.88
General Farming	699	710	109.40	19.19	1.03	0.18	0.01	0.91
Hay	609	4,640	128.11	14.94	.55	0.12	0.02	1.33
Vegetables	557	1,850	132.74	36.92	2.03	0.28	0.02	0.93
Fruit or Nut Orchard	253	9,830	442.65	99.08	1.72	0.22	0.03	0.98

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Endrin							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	2,130	142.72	1.73	.05	0.01	0.00	0.81
Corn	1386	80	37.11	.15	.01	0.00	0.00	0.97
Wheat & Small Grains	1056	80	30.76	.34	.04	0.01	0.00	0.74
Cotton	221	420	111.21	7.66	.32	0.07	0.02	0.87
Soybeans	1271	640	93.08	.88	.03	0.01	0.00	0.87
General Farming	699	480	234.28	.72	.01	0.00	0.00	1.08
Hay	609	100	58.40	.16	.01	0.00	0.00	0.83
Vegetables	557	1,000	187.72	6.91	.17	0.04	0.01	0.77
Fruit or Nut Orchard	253	2,130	483.73	13.92	.15	0.03	0.01	1.05

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Heptachlor							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	1,710	101.01	4.78	.20	0.05	0.00	0.81
Corn	1386	1,710	112.48	12.21	.51	0.11	0.01	0.84
Wheat & Small Grains	1056	10	10.00	.02	0.00	0.00	0.00	0.81
Cotton	221	10	10.00	.10	.02	0.01	0.01	1.07
Soybeans	1271	940	102.72	9.57	.42	0.09	0.01	0.95
General Farming	699	290	47.69	.99	.07	0.02	0.01	1.04
Hay	609	260	56.48	.34	.02	0.01	0.00	0.76
Vegetables	557	30	16.74	.18	.03	0.01	0.00	0.92
Fruit or Nut Orchard	253	190	75.22	1.23	.07	0.02	0.01	1.06

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Heptachlor Epoxide							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	1,080	54.59	4.24	.31	0.08	0.00	0.92
Corn	1386	350	54.75	9.28	.82	0.17	0.01	0.93
Wheat & Small Grains	1056	70	24.69	.17	.02	0.01	0.00	0.82
Cotton	221	40	21.50	.62	.09	0.03	0.01	1.10
Soybeans	1271	1,080	64.00	7.56	.54	0.12	0.01	0.98
General Farming	699	200	38.18	1.59	.15	0.04	0.01	0.96
Hay	609	720	68.05	2.17	.12	0.03	0.01	1.32
Vegetables	557	120	30.32	1.37	.15	0.05	0.01	1.45
Fruit or Nut Orchard	253	180	44.25	2.88	.23	0.07	0.02	1.06

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Isodrin							
	n[1/]	Max[2/]	$\bar{x}_+$[3/]	$\bar{x}$[4/]	$\bar{x}_g$[5/]	P(>MDL)[6/]	S.D.[7/]	DEFF[8/]
Total RSN	6071	180	21.68	.16	.02	0.01	0.00	0.96
Corn	1386	180	21.46	.47	.06	0.02	0.00	0.96
Wheat & Small Grains	1056	0	0	0	0	0.0	0.0	1.00
Cotton	221	0	0	0	0	0.0	0.0	1.00
Soybeans	1271	90	24.23	.27	.03	0.01	0.00	1.10
General Farming	699	20	14.98	.04	.01	0.00	0.00	1.04
Hay	609	0	0	0	0	0.0	0.0	1.00
Vegetables	557	10	10.00	.02	0	0.00	0.00	1.13
Fruit or Nut Orchard	253	0	0	0	0	0.0	0.0	1.00

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Toxaphene							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	36,330	3,562.56	129.98	.32	0.04	0.00	0.64
Corn	1386	8,800	2,761.23	18.74	.05	0.01	0.00	0.77
Wheat & Small Grains	1056	1,600	810.18	1.78	.01	0.00	0.00	0.78
Cotton	221	36,330	4,190.03	1,394.85	11.81	0.33	0.03	0.88
Soybeans	1271	21,000	3,932.77	261.63	.67	0.07	0.01	0.65
General Farming	699	2,080	2,080.00	2.99	.01	0.00	0.00	1.00
Hay	609	11,030	5,174.00	22.66	.04	0.00	0.00	0.89
Vegetables	557	12,000	2,734.31	216.55	.82	0.08	0.01	0.91
Fruit or Nut Orchard	253	8,300	2,601.99	267.90	1.16	0.10	0.02	0.89

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Trifluralin							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	6071	1,860	99.33	3.20	.14	0.03	0.00	0.80
Corn	1386	600	88.20	2.79	.14	0.03	0.00	0.93
Wheat & Small Grains	1056	290	126.32	.36	.01	0.00	0.00	0.76
Cotton	221	310	70.07	11.12	.86	0.16	0.02	0.82
Soybeans	1271	680	87.60	6.41	.35	0.07	0.01	0.88
General Farming	699	310	104.94	.87	.03	0.01	0.00	0.99
Hay	609	10	10.00	.01	0.00	0.00	0.00	0.86
Vegetables	557	1,860	160.00	7.40	.22	0.05	0.01	0.94
Fruit or Nut Orchard	253	1,290	328.65	5.33	.06	0.02	0.01	1.02

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Arsenic							
	n[1]	Max[2]	$\bar{x}_t$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	4690	180,420	5,869.29	5,665.15	2,863.07	0.97	0.00	1.44
Corn	1109	31,980	5,653.68	5,467.48	3,101.61	0.97	0.01	1.38
Wheat & Small Grains	826	37,530	5,292.43	5,091.67	2,616.57	0.96	0.01	0.81
Cotton	177	38,900	5,932.09	5,827.79	3,497.19	0.98	0.01	1.05
Soybeans	962	107,450	6,723.70	6,521.48	3,462.38	0.97	0.01	1.01
General Farming	516	64,940	6,376.99	6,234.96	3,360.02	0.98	0.01	1.05
Hay	453	51,300	5,602.92	5,275.52	2,058.05	0.94	0.02	2.84
Vegetables	448	69,100	4,997.11	4,851.67	2,367.47	0.97	0.01	1.01
Fruit or Nut Orchard	182	180,420	8,009.47	7,654.21	2,415.32	0.96	0.02	1.04

(continued)

Table 1.7.9: Statistics for Compounds with Detectable Levels in Cropland Soils by Cropping Region for Round One (continued)

Cropping Region	Atrazine							
	n[1]	Max[2]	$\bar{x}_+$[3]	$\bar{x}$[4]	$\bar{x}_g$[5]	P(>MDL)[6]	S.D.[7]	DEFF[8]
Total RSN	523	16,730	231.40	115.34	8.30		0.50	0.02
Corn	271	1,550	185.62	93.25	9.18	0.50	0.03	1.18
Wheat & Small Grains	26	120	94.17	17.73	1.32	0.19	0.12	2.46
Cotton	0	-	-	-	-	-	-	-
Soybeans	102	16,730	537.13	284.52	10.59	0.53	0.05	0.98
General Farming	89	1,380	113.52	62.55	8.68	0.55	0.05	1.04
Hay	17	100	43.80	15.00	2.46	0.34	0.12	1.11
Vegetables	16	340	110.04	82.47	20.33	0.75	0.10	0.87
Fruit or Nut Orchard	1	40	40.00	40.00	39.85	1.00	0.0	1.00

(continued)

[1] Sample size.

[2] Maximum amount detected (ppm).

[3] Weighted average of the data values in excess of the MDL (ppm).

[4] Weighted average of the amount detected (ppm).

[5] Antilog$_e$ (weighted average of log$_e$ (amount +1)-1); analogous to the geometric mean (ppm).

[6] Weighted proportion of cases with data values in excess of the MDL.

[7] Standard deviation of the estimated proportion.

[8] Design effect for the estimated proportion.

* Source: Computer files supplied by the EPA Field Studies Branch, Washington, D.C.

undoubtedly conservative estimates. Thus, the interval of values within two standard errors of the estimated proportions will provide a conservative 95 percent confidence interval estimate of the proportion of sampled area where levels of the compound exceed the minimum detectable level (MDL).

The design effect is the ratio of the sample standard error to an estimate of what the standard error would have been if a simple random sample of the same size had been used, i.e.

$$\text{DEFF} = \frac{\text{Estimated S.E. (For the design used)}}{\text{Estimated S.E. (Simple Random Sample)}}$$

Alternatively, the design effect can be thought of as the ratio of the actual sample size to the sample size that would be required to obtain an estimate with the same standard error based upon a simple random sample. Generally stratification decreases the design effect, while clustering increases it. Thus, since the CNI stratification can be used and there is no clustering of sample sites in the RSN sample, design effects less one would be expected. This would indicate that the design produced smaller standard errors than would a simple random sample of the same size. Many of the design effects shown in Tables 1.7.7 through 1.7.9 are indeed less than one. However, some design effects are substantially greater than one. It is, hence, not clear that the CNI stratification was particularly advantageous for estimation of proportions of detections for toxic substance residues.

1.8 Capabilities for Performing Special Studies

If it were possible to completely fulfill the design of the Rural Soils Network (RSN), it would serve as an excellent vehicle for performing special studies. With one-fourth of all sites in each State being sampled in each year, baseline levels of pesticide residue would soon be established for all moderate size geographic areas. Data needed for special studies of specific pesticides or specific areas would then be readily available.

1.9 Toxic Substances Other Than Pesticides in Soils

The NSMP currently monitors three classes of pesticides in soil, organchlorine pesticides and trifluralin, organophosphorus pesticides, and heavy metals. Each of these classes are analyzed using methodology specifically designed to provide optimum selectivity and sensitivity for that class to the exclusion of others.

Expanding the capability of the soil networks in monitoring for a wide range of toxic substances will require the development of analytical methodology to deal with the special characteristics of these substances as well as those of the matrix. A wide range of new techniques (e.g., high performance liquid chromatography, mass spectroscopy, electrochemistry and capillary gas chromatography) may need to be incorporated to accomplish this purpose. However, the design and application of effectively administered QC/QA programs must be concurrent with the development of appropriate analytical methodology.

Much of the necessary methodology is already available in the open literature or in EPA and contracting laboratories. Some may be directly applicable to the perceived needs of the NSMP, and others will require some degree of modification to account for differences in either the analyte or matrix. All aspects of the methodology must be evaluated (i.e., sample collection and storage, analyte isolations and instrumental analysis) and the method appropriately validated in order for the NSMP to meet the needs of those who are using the analytical data.

The working definition of "toxic substance" at present must include virtually any substance manufactured in or imported into the United States. Great care must be exercised in decisions regarding the choice of substances to be monitored by NSMP. The complexity and cost of the required methodology increase directly with the number of substances and matrices to be analyzed. Thus, misjudgement can quickly lead to unnecessary or nonproductive expenditures of time and funds.

There are two major methodological approaches to the concurrent analysis of a number of different substances. The first approach is the development of a "survey" method in which specimen components are separated only to the extent necessary to ensure the compatability of each component with the analytical technique. The resulting subset of specimens are all for the analysis of such specimens; however, the overall number of specimens requiring analysis is minimized. Two such "survey" methods (Master Scheme for the Analysis of Organic Compounds in Water and A Comprehensive Method for the Analysis of Volatile Organisms on Solids, Sediments and Sludges) are currently being developed under EPA. Development of a truly all-inclusive "survey" method may be neither possible nor practical as the present methods are limited to analysis of organic compounds which are or can be made sufficiently volatile to pass through a capillary gas chromatograph/mass spectrometer.

An alternate approach is the development of analytical methods optimized for a specific substance or class of substances. Each method necessarily excludes all substances except those of similar chemical and physical characteristics. Monitoring of a large number of different substances would therefore require the use of a number of specific analytical methods.

Neither approach is without its disadvantages and these must be weighted against the goals of the monitoring network. A basic philosophy must be established regarding these goals and the methodology approach which will best serve them over the long term.

1.10 Implementation Plan for a New Survey Design of the Rural Soils Network

A specific implementation cannot be recommended at this time, since a specific design option has not been recommended. One observation which can be made is that the transition period should cover one cycle in the old design, 4 years. Since, it is not likely to be feasible to investigate the entire RSN, nor indeed is it necessary, a subset of the old sites may be used. An advantageous scheme may be to link old and new sites on the basis of geographic proximity, and compare their observations over the transition period.

2. EVALUATION OF CHEMICAL ANALYSIS

2.1 Objective

The objective of this section is to conduct a limited review of the current analytical methodology used in the National Soils Monitoring Program (NSMP) in order to assess the quality and reliability of the data with respect to meaningful statistical evaluation and statistical survey design.

2.2 Discussion

Data compiled by NSMP is generated by the use of complex multiresidue analytical methodology, and the quality of such data is determined primarily by the limitations of the methodology. These limitations are normally defined in terms of the precision, accuracy and minimum detectable level (MDL) of the analytical method for each specific analyte and specimen matrix. A knowledge of these limitations is especially important to potential users of the analytical data since reported substance levels are merely estimates of the "actual" levels in the matrix. As estimates, individual values in the NSMP data file in fact represent ranges (of values) in which the "actual" substance levels are reasonably (i.e., with some high probability) expected to fall. The size of the range can be adequately described by the accuracy and precision of the analytical method; therefore, a knowledge of these parameters is required for meaningful evaluation of the data. In addition, the MDL for the method defines (or should define) the lowest level that can be estimated with reasonable confidence, no analytical method being capable of absolute detection down to zero concentrations. This limit must be considered in evaluating the practical versus the statistical significance of trace levels and zeros reported in the data file (Hartwell et al. 1979).

The extensive manual of recommended analytical methodology has been published by EPA-RTP (USEPA. 1977) for use in routine multiresidue pesticide analysis. The complexity of the sample matrices and pesticide types routinely analyzed in practice requires that the methodology consist of a basic analytical procedure with a large number of modifications and ancillary techniques in order to cope with problems imposed by widely divergent pesticide levels and interferences. Each modification or technique produces a specific effect on the accuracy, precision and MDLs of the overall analytical method, and, hence, must be validated for each pesticide analyzed by the method. A detailed knowledge of the analytical procedure is therefore required in order to properly assess the quality of the data generated by the procedure.

An extensive set of recommended QC/QA procedures has been published by EPA-RTP (USEPA. 1979) in an effort to control the quality of data produced by analysts and laboratories using the multiresidue pesticide method. Laboratories adhering to these recommendations will necessarily generate (through controls, blanks, and SPRMs*) much of the information needed to assess accuracy, precision and MDLs for data reported to the NSMP. Control and SPRM data are not, however, compiled or summarized in

*Special Pesticide Refrence Material.

a single document (i.e., issued semiannually or annually), or entered into the computer data file. Thus, for all practical purposes, the data are lost to the potential data file users. Reporting of all control data in the computer file, along with results for soil specimens, would allow the data quality to be determined according to the specific needs of the individual user (e.g., for a particular pesticide in a given geographic area or over a specified period of time). The results of duplicate specimen analysis are apparently not reported in the computer data file. Again, this is valuable information lost to computer file users.

RTI has attempted to review the analytical methodology used to generate data under the NSMP and compile existing information on the current quality of the data (accuracy, precision and MDL) in order to make this information available to Program data file users. The review is necessarily limited by the provisions outlined in the revised work plan for this task.

In the interest of clarity and accuracy, the RTI request for detailed information on analytical procedures and data quality was made in written form. A questionnaire was submitted to William G. Mitchell of the Toxicant Analysis Center, Bay St. Louis, Mississippi, the laboratory currently responsible for carrying out chemical analysis under the NSMP. The cover letter to Mr. Mitchell and the questionnaire are given in Appendix A. The questions were designed to provide detailed information on all areas of current analytical methodology pertinent to the quality of the data generated by the method. It was anticipated that extensive verbal follow-up (telephone) would be required to obtain additional information and clarify details. The initial information from the laboratory has been received by RTI and evaluation of the information carried out. The results are presented below.

2.2.1 Analytical Methodology

The NSMP currently reports levels in soil for over thirty pesticides and toxic substances (Table 2.1) including several chemical classes (i.e., organochlorine and organophosphorous pesticides, trifluralin and heavy metals). All analyses are carried out at the Toxicant Analysis Center (TAC) in Bay St. Louis, Mississippi. The analytical results for each soil specimen are reported on a single form (Appendix B) along with the specific location and date at which the specimen was taken. Individual pesticides and metals detected in the specimen are listed along with their levels in fourteen blank spaces on the form. Reporting units (i.e., ppm, ppb & ppt) are specified using a value code following the particular result. Although there are spaces on the form for individual soil characteristics such as pH, % sand, % silt, % clay and % organic matter; these characteristics are not currently determined for urban soil specimens. The % moisture content of each soil specimen is determined but not reported on the form. The reported results are, however, corrected for % moisture (i.e., reported on the basis of dry solid weight). An important point of confusion arises from the use of the term chlordane in reporting results. The term usually corresponds specifically to the level of γ-chlordane in a soil specimen as this is the most commonly found isomer. However, when the α-isomer and t-nonachlor

Table 2.1. Pesticides and Toxic Compounds Analyzed Under NSMP

Organochlorines	Organophosphates	Heavy Metals
Alachlor	DEF	Mercury
Aldrin	Diazinon	Cadmium
BHC	Ethion	Lead
Chlordane	Malathion	Arsenic
DDTs	Phorate	
Dieldrin	Parathion, ethyl	
DCPA	Parathion, methyl	
Dicofol	Ronnel	
Endosulfan I	Trithion	
Endosulfan II		
Endosulfan Sulfate		
Endrin		
Endrin Ketone	Other	
Heptachlor	Trifluralin	
Heptachlor Epoxide		
Hexachlorobenzene		
Isodrin		
Lindane		
Methoxychlor		
PCBs		
Propachlor		
Toxaphene		

Source: Toxicant Analysis Center (TAC), USEPA, Bay St. Louis, Mississippi.

(and presumably oxychlordane since it is not listed separately in Table 2.1) are also found, all levels are reported under the term chlordane as their sum. The term "technical chlordane" is inappropriate as one of the major components, heptachlor, is reported separately. In order to avoid confusion in the subsequent interpretation of the data, all individual components of pesticide mixtures (e.g., chlordane, BHC and PCB) should be reported as such. Otherwise potentially valuable data is lost.

Only urban soil specimens are currently collected and analyzed at TAC; the last rural soil specimens having been analyzed in 1977. The soil specimens are collected by either EPA or a contracting laboratory as a pattern of 6-8 core specimens composited in a one quart, wide-mouth, glass mason jar with a Teflon- or aluminum foil-lined cap. Specimens are subsequently shipped to TAC at ambient temperature. Specimens received at TAC are refrigerated until they can be analyzed. The specimen collection date, the date of receipt at TAC, and the date of analysis are all stated on the result form, and thus, these data are presumably available to computer data file users.

The analytical methodology used in the analysis of organochlorine and organophosphorous pesticides in soil specimens is essentially the same as that used for the analysis of sediment specimens under the National Surface Water Monitoring Program (D. Lucas et al. 1980). The analysis of pesticides in both matrices is performed at TAC. The specific procedure for the extraction and Florisil clean-up of soil specimens for analysis of organochlorine and organophosphorous pesticides is given in Appendix C. The procedure was furnished by TAC as a result of the RTI questionnaire. The levels of pesticides in specimen extracts are determined using essentially the same gas chromatographic techniques applied to water and sediment (D. Lucas et al. 1980). Although trifluralin is a nitroaniline, its chemical properties allow it to be analyzed with the organochlorine pesticides.

The general method used in the analysis of organochlorine and organophosphorous pesticides involves an initial screening of specimen extracts on a primary GC system. All positive results are then screened on a secondary GC system that differs from the primary system in the selectivity of either the column or the detector. Continued positive results <u>may</u> then be confirmed through the use of additional analytical techniques depending on the degree of suspected difficulties from interference, contamination, or low levels (approaching the MDL). The techniques used in the application of this methodology to each pesticide group are summarized in Table 2.2.

Quantitation of GC results for pesticides is carried out using external standard procedures with single-point calibration. Calibration standard concentrations are adjusted to give compound responses similar to those in specimen extracts in order to reduce the effects of detector non-linearity. The ECDs used in this program all possess an ^{63}Ni source.

Actual recoveries of pesticides from soil specimens are not monitored except <u>via</u> the corresponding recoveries for controls. It would be extremely useful to fortify each specimen with a particular compound(s)

Table 2.2. Procedures for the GC Analysis of Pesticides for the NSMP

Compound Class	Primary analysis	Secondary analysis	Confirmation techniques	Additional comments
Organochlorine	GC/ECD on OV-1	GC/ECD on OV-210	GC/HECD GC/ECD on 1.5% OV-1/ 1.95% OV-210	Every 10th soil analysis duplicated
Organophosphorus	GC/FPD on OV-1	GC/FPD on 1.5% OV-1/ 1.95% OV-210	GC/NPD GC/FPD-S mode GC/ECD	Every 10th soil analysis duplicated

Source: Toxicant Analysis Center (TAC), USEPA, Bay St. Louis, Mississippi.

prior to extraction and clean-up and thereby monitor any anomalous behavior in the extraction, clean-up and GC injection procedures which may occur from time to time for a particular specimen. This technique was used in the National Human Monitoring Program for analysis of organochlorine pesticides in adipose tissue. Aldrin, which is seldom found, was spiked into fat specimens and then analyzed as though it were endogenous. The internal standard quantitation method is an alternate procedure for normalizing recoveries and was briefly examined at Versar, Inc. for the analysis of s-triazines in water and sediment specimens (D. Lucas et al. 1980). Promazine was used as the internal standard and preliminary work showed promise (Bob Martin, Versar, Inc.).

All GC methodology used in the NSMP utilizes packed-column techniques. The improved resolution and sensitivity that could be obtained by incorporating state-of-the-art capillary GC techniques would considerably increase the utility of the method and reduce the need for confirmation. This is particularly evident in the analysis of PCBs where each individual designation (i.e., Arochlor 1242, Arochlor 1254 and Arochlor 1260) actually represents a complex mixture of partially chlorinated biphenyls (e.g., tri-, tetra-, penta- and hexachlorobiphenyl) and their respective isomers. Considerable overlap exists between the components present in each PCB. For instance, Arochlor 1242 contains di- to hexachlorinated biphenyls whereas Arochlor 1254 contains tetra- to heptachlorinated biphenyls. There is currently no reason to expect the individual components to possess the same degree of stability or toxicity. Thus, the original pattern of components and their relative amounts may not be preserved in complex environmental and biological matrices. Yet, it is on the basis of the standard peak pattern that the presence of PCBs and their levels are currently determined. Further, as was shown in the analysis of fat specimens under the National Human Monitoring Program (R.M. Lucas et al. 1980), PCB components can interfere with the analysis of some chlorinated pesticides (e.g., p,p´-DDT, t-nonachlor and heptachlor epoxide) at sufficiently high levels. The degree of resolution that can now be achieved by capillary GC techniques is demonstrated by the chromatogram of Arochlor 1242 and 1260 in Figure 2.1. These Arochlors cover nearly the entire range of PCB components (monochlorobiphenyls to octachloropiphenyls and their isomers) and yield over 80 individual peaks by this method. Typical packed-column GC techniques yield less than 15 peaks for these mixtures.

In general, the analysis of heavy metals in soil specimens was carried out using flame atomic absorption (AA) techniques for cadmium, lead and arsenic(T.J. Forehand et at, 1976), and the cold vapor AA technique for mercury. More specific information on the current methodology was unavailable for two reasons. First, soil specimens have not been analyzed for heavy metals since 1979. In view of recent instrumental acquisitions (i.e., graphite furnace and Zeeman AA) coupled with the continual refinement of AA procedures in the ongoing analysis of other matrices (e.g., blood, urine, etc.) at TAC; it is likely that the original procedures (for soil) will undergo substantial modification when the analysis of soil specimens is resumed. Second, the individual responsible for the most recent analysis of soil specimens (1979) is no longer employed at TAC, and thus detailed information concerning the methodology and control data is not readily available. It has been necessary to

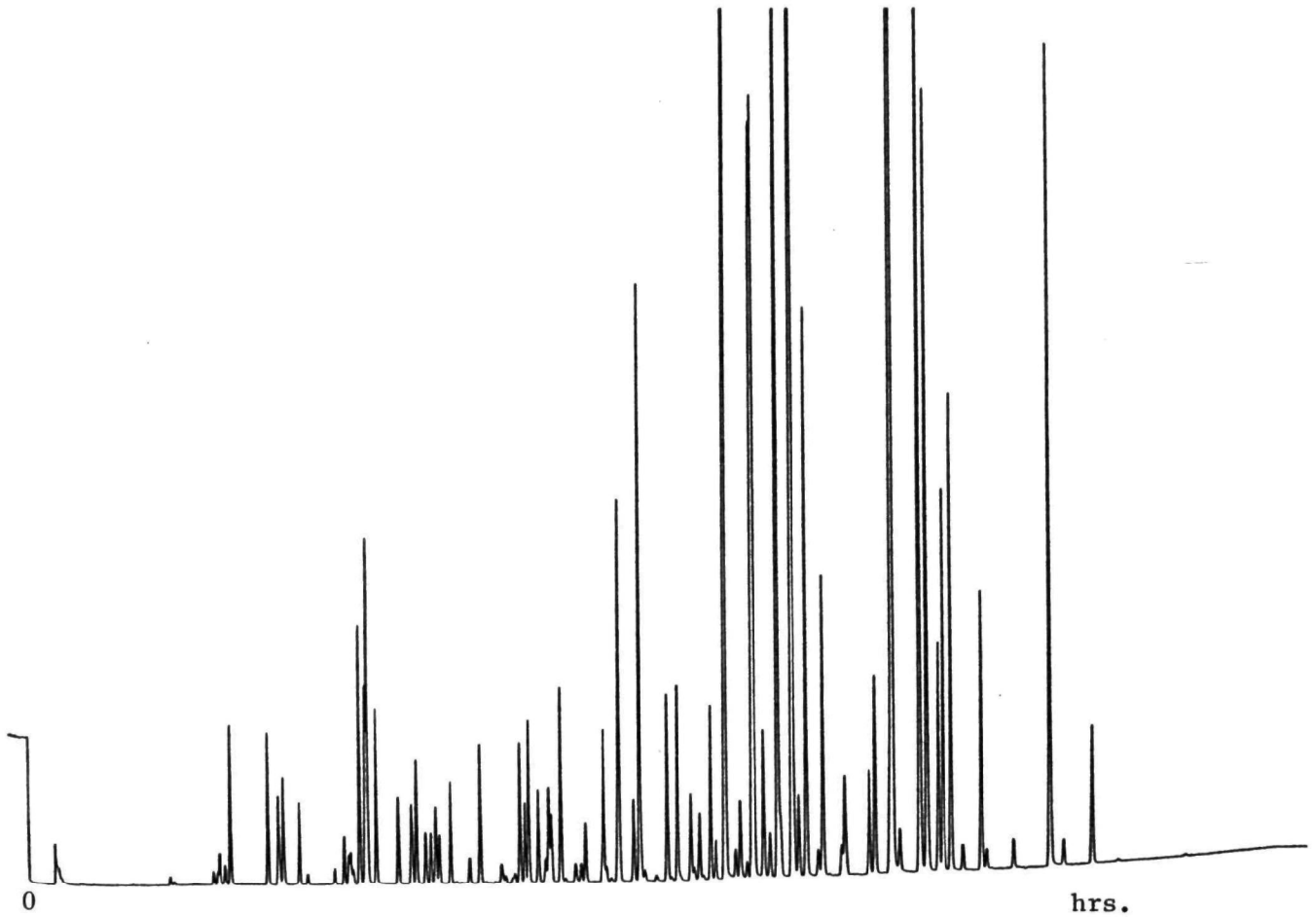


Figure 2.1 Capillary GC/ECD Chromatogram of Arochlor 1242 and Arochlor 1260 (∿12 ng total): 48m x 0.25mm id capillary with 0.1μ Apeizon L on persilylated pyrex, 1.5mL/min. helium, 150°-290° @1°/min., ECD @ 128 x 10.

Source: Toxicant Analysis Center (TAC), USEPA, Bay St. Louis, Mississippi.

obtain such information directly from the analyst for virtually every analytical method used in the five National Monitoring Programs; a situation which further demonstrates the need for centrally located documentation of all methodology used in a monitoring program. In view of the lack of available data on the current analysis of metals in soil, the quality of the analytical data cannot be determined at this time.

2.2.2 QC/QA

For organochlorine-organophosphorus pesticides in soil, individual specimens are analyzed in sets of 10-15 with each set containing a method blank (reagent blank) and a control. The control consists of a fortified soil specimen (SPRM), which is generated internally. Checks are also run on the elution pattern of pesticides from Florisil columns.

Information regarding the primary and secondary analytical techniques and confirmation techniques that may be used in the analysis of soil specimens has been summarized in Table 2.2. Decisions concerning the adequacy of the clean-up procedure (if used), the validity of the standards and controls, and the confirmation techniques used are reviewed by the supervisor (i.e., William Mitchell) and the TAC QC/QA officer (Dr. Joe Yonan).

2.2.3 Accuracy and Precision

Information on the accuracy and precision of an analytical method is required in order to define the relationship between the analytical result (estimate) and the "actual" analyte level in the specimen. Although this information may be produced as part of the initial method development and validation, it is by itself insufficient as the characteristics of the method can (and frequently do) change with time, analyte level, and matrix. Environmental matrices are particularly complex and variable.

The replicate analysis of a specimen containing a known level of analyte (i.e., a control) over a period of time can provide useful information about the accuracy and precision of the method, and the method stability. RTI has attempted to compile such information, where it is available, for each pesticide and toxic substance listed in Table 2.1.

The accuracy of the analytical methodology is probably best reflected in the recovery of the analyte. This is particularly true for the toxic substances monitored in the NSMP since the analytical results are not corrected for losses during workup (i.e., recoveries). The available recovery data for analysis of toxic substances in soil is given in Table 2.3. These values represent averages over a period of several months and are therefore more useful as general indications of method accuracy than a corresponding single value. Unfortunately, this information is far from complete with respect to the number of toxic substances listed (Table 2.3) versus the number analyzed (Table 2.1). Recovery data must be generated for all substances analyzed. While a single value may be found to hold for a number of similar substances,

Table 2.3. Average Recoveries for Some Organochlorine Pesticides from Soil

Pesticide	Fortification level (ppb)	Average % recovery	Average error	Reported MDL (ppb)
γ-Chlordane	60	80	-20%	10
o,p'-DDE	90	81	-19%	20
p,p'-DDE	75	84	-16%	20
p,p'-DDD	150	87	-13%	20
o,p'-DDT	240	84	-16%	20
p,p'-DDT	240	88	-12%	20
Dieldrin	90	80	-20%	20
Aldrin	30	127	+27%	10
Heptachlor	30	80	-20%	10
Heptachlor Epoxide	60	80	-20%	10
Endrin	120	89	-11%	20

Source: Toxicant Analysis Center (TAC), USEPA, Bay St. Louis, Mississippi.

the similarity must be demonstrated (and the substances thus grouped must be specified) and does not obviate the need for subsequent monitoring via controls.

Available information on the analytical precision for toxic substances in soil is given in Table 2.4 in terms of the coefficient of variation (CV). The CV is calculated as follows:

$$CV = \frac{\text{std. deviation}}{\text{mean}} \times 100 = \% \text{ relative standard deviation.}$$

As with recoveries, data was available for only a small number of organochlorine pesticides and the fortification levels were variable (3-12 times the MDL).

No indication of specific interferences between pesticides has been given. This is particularly interesting since PCB levels greater than 1 ppm were found to significantly interfere in the GC/ECD analysis of p, p'-DDT, t-nonachlor and heptachlor epoxide in human fat (RM Lucas et al. 1980). The GC methodology used for the analysis of organochlorine pesticides in water and sediment does not appear to differ significantly from that for human adipose tissue. Thus the ubiquitous nature of PCBs would be expected to cause interference problems regardless of the specimen matrix. The use of high resolution capillary GC techniques would contribute significantly to the elimination of such difficulties, as well as increase the sensitivity of GC/MS as a highly specific confirmation procedure.

2.2.4 Minimum Detectable Levels

All analytical techniques are characterized by an inherent limit of sensitivity below which the technique cannot reliably discern the presence or absence of a particular component. Thus the procedures used in the analysis of soil specimens must be similarly characterized by a minimum amount of specific analytes which produce a signal response statistically discernible from background. This analyte concentration is defined as the minimum detectable level (MDL) and is important in the assessment of the analytical data since concentrations reported below the MDLs lack validity and must be considered unreliable.

The MDLs associated with the GC analysis of specific pesticides in soil specimens are a function of instrumental operating conditions and the amount of background introduced by residual matrix material in the injected specimen extract. Tentative detection limits have been established by TAC and are shown in Table 2.5 for the organochlorine and organophosphorous pesticides.

The MDL corresponds to the amount of analyte producing a signal equal to 5% of full scale deflection with a maximum of 1% noise (signal to noise ratio = 5:1). In cases where the chromatographic background is significant, the MDL is taken as that amount of analyte producing a signal equal to twice the noise level in the vicinity of the peak.

Table 2.4. Precision for Some Organochlorine Pesticides in Soil

Pesticide	Fortification level (ppb)	CV
γ-Chlordane	60	3%
o,p'-DDE	90	2%
p,p'-DDE	75	3%
p,p'-DDD	150	3%
o,p'-DDT	240	2%
p,p'-DDT	240	4%
Dieldrin	90	2%
Aldrin	30	10%
Heptachlor	30	3%
Heptachlor Epoxide (HE)	60	4%
Endrin	120	5%

Source: Toxicant Analysis Center (TAC), USEPA, Bay St. Louis, Mississippi.

Table 2.5. Detection Limits of Pesticides in Soils

Compound	Detection limit
Organochlorine pesticides	
Early eluting (BHC's, Aldrin, Heptachlor Epoxide, Chlordane)	10 ppb
Late eluting (DDTs, Dieldrin, Endrin)	20 ppb
All multicomponent pesticides	50 ppb
Organophophorous pesticides	10-50 ppb

Source: Toxicant Analysis Center (TAC), USEPA, Bay St. Louis, Mississippi.

Any response lower than the MDL is reported as not detected (ND). There can be significant variation in analytical sensitivity, even among specimens of the same type. Consequently, the reported MDLs are typical or expected levels realized for the majority (75-80%) of specimens.

2.3 Fate of Pesticides in Soil

After the application of pesticides to agricultural land a number of processes may occur which lead to its transport in the environment or its removal by chemical or biological degradation. Both of these mechanisms depends to a large extent on the chemical structure of the pesticide and to a lesser extent on the soil type, clay, clay loam, sandy loam, sand etc.

Chlorinated hydrocarbons have a well earned reputation for persistence. Kearney, Nash and Isensee has compared persistence of pesticides within each general pesticide type and gives the persistence of chlorinated hydrocarbons varies from 5 years for chlordane to 2 years for heptachlor and aldrin. The persistence of phosphate insecticides is measured in weeks by contrast. Diazinon persists for some 12 weeks compared to Malathion and Parathion which persist for only a few weeks. Trichloroacetic acid persists for 12 weeks compared to 2 weeks for Barban. Intermediate between the two extremes are a wide range of herbicides. The urea, triazine and pieloram herbicides range from 3 months for Prometryne to 18 months for Picloram and Propazine. The benezoic acid and amide herbicides range from 2 to 12 months and the phenoxy, toluidine and nitrile herbicides range from 1 to 6 months persistence.

The migration of pesticides in soils is again very closely related to its chemical structure. Such factors as water solubility and the absorption on soil particles affect their migration compounds such as Aldecarb have migrated through the soil over burdened to shallow aquifers. Halogenated hydrocarbon pesticides such as benezene hexachloride which has a low water solubility remains entirely in the upper soil layer (2).

The effect of soil type on the fate of pesticides has been studied by a number of workers. In very general terms absorption is greater on clays than on sand. The organic content of the soil also affects adsorption (3).

(1) Kearney, P. C., R. G. Nash and A. R. Isensee 1979. Persistence of Pesticide Residues in Soils. In M. W. Miller and G. G. Berg (ed). Chemical Fallout, Current research on persistent pesticides, Charles C. Thomas, Publisher, Springfield, Ill.

(2) Kawahara, T. M., Matsui and H. Nakamura, "BHC in Soil of Paddy Field" Bull. Agric. Chem. Inspec. Stn. 12:42-45 (1972).

(3) Bristow, P. R., J. Katan and J. L. Lockwood. "Control of Rhizoctoria solani by Pentachloronitrobenzene Accumulated from Soil by Bean Plants," Phytopathology 63:808-813(1973).

2.4 Recommendations

Limited review of analytical methodology used in the NSMP and an attempt to compile data for the average accuracy, precision, and MDL in soil for each toxic substance monitored under this program provide a basis for the following recommendations:

1. Accuracy (that is, recoveries) and precision data must be generated for all pesticides monitored in the NSMP. The data should be generated at two different levels (e.g., at the MDL and at ten times the MDL). The results for controls analyzed with each set of specimens would be the best means of providing this information since it is necessary that control data be made accessible to computer data file users in any event. Controls must be run with each set of specimens and should consist of a blank (unfortified soil free from the analytes of interest) and two fortified blanks (one fortified at the MDL and another at ten times the MDL). The analytical results for the controls should be reported on a separate form (especially designed for control data) and encoded such that there is a one-to-one association with the particular set of specimens with which they were analyzed. The encoding should allow later computer retrieval of control data for any particular specimen set or group of sets (for example, geographic area, over a specified period of time, or for a particular pesticide). The availability of this information in a retrievable form to data file users would provide the means for assessing data reliability now lacking. Further, any duplicate specimen analyses must be reported in the computer data file as they provide the best means of assessing method precision on a continuous basis. Duplicate results must be specifically encoded such that they are retrieved as a group (e.g., all duplicates for a particular matrix and pesticide over a specified period of time) as well as with the initial analytical results for the specimen. The need to make routine control data available to program data file users cannot be overemphasized. This does not preclude the use of specialized controls (e.g., SPRMS,); however, these results should also be included in the computer file encoded to allow facile retrieval both as a group and with their particular specimen set.

2. The pesticides included on the routine monitoring list must be reviewed on a regular basis and appropriate deletions or additions made. Specifically, the need for routine analysis of organophosphorous pesticides in soil should be reviewed as this class of compounds is known to be unstable and has seldom been reported in either soil or sediment. Once the baseline has been established for such compunds, three choices are possible: 1) cease to analyze for the compound(s) except under special circumstances (e.g., after a chemical spill or when contamination is suspected from a recent application); 2) analyze for the compound(s) on a more infrequent basis; and 3) concentrate efforts on the analysis of degradation products of known toxicity where these exist. Decisions concerning the

analysis of toxic substances under the NSMP should be based on information generated in other agency data files (e.g., USDA, USGS, etc.) as well as data generated within EPA.

3. Soil specimens should be characterized as to the percent carbon or percent inogranic residue. This information must be included on the report form (along with moisture content) as part of the specimen characterization (source). Significant trends may otherwise be missed with respect to the soil type and its effects on toxic substance accumulation, degradation and transport.

4. Control specimens (in the matrix of interest) should be included with any specimens either stored for extended periods or shipped to another site for analysis. This is particularly important for toxic compounds which are known to be unstable; i.e., organophophorous pesticides. The results of these "storage controls" must also be included in the computer data file with appropriate encoding for specific retrieval.

5. Analytical methodology should be updated to include state-of-the-art capillary GC techniques. This would provide a higher degree of confidence in the resuling data through increased resolution and sensitivity. The use of higher resolution analytical techniques is a move toward the quantitation of PCBs (and technical chlordane) as their individual isomers. This approach is far more useful than the present method of attempting to identify patterns and averaging components, since the toxicity and biodegradation of the individual isomers are not identical.

6. The pesticide recoveries should be monitored for each specimen analyzed by initial fortification of the specimen with appropriate compound(s). Subsequent analysis of the compound level should enable comparison of data between specimens with increased confidence that anomalous results will be detected. The use of internal standard quantitation techniques would normalize recoveries between specimens and should be considered.

7. Detailed information on <u>all</u> analytical procdures under under the NSMP should be documented in one source. The procedures must then be maintained current with ongoing improvements and modifications made by the analytical laboratories. Such updating requires both flexibility and regular review by program management.

References

(Analytical Section)

Hartwell TD, Piserchia P, White SB et al. 1979. Analysis of EPA Pesticides Monitoring Networks. Washington, DC: U.S. Environmental Protection Agency. EPA-560/13-79-014.

USEPA. 1977. U.S. Environmental Protection Agency. Manual of Analytical Methods for the Analysis of Pesticide Residues in Human and Environmental Samples. Revised June 1977 under EPA Contract No. 68-02-2474.

USEPA. 1979. U.S. Environmental Protection Agency. Manual for Analytical Quality Control for Pesticides and Related Compounds in Human and Environmental Samples. Revised January 1979 under EPA Contract No. 68-02-2474.

Lucas D, Mason RE, Rosenzweig M et al. 1980. Recommendations for the National Surface Water Monitoring Program: Report Two. Research Triangle Park, NC: Research Triangle Institute. RTI/1864/14/01-02I.

Lucas RM, Rosenzweig MS, William SR et al. 1980. Evaluation of and Alternate Designs for National Human Monitoring Program's Adipose Tissue Survey. Research Triangle Park, NC: Research Triangle Institute. RTI/1864/14/02-2I.

Forehand TJ, Dupuy AE, Tai H. 1976. Determination of arsenic in sandy soils. Analytical Chemistry 48(7): 999-1001.

APPENDIX A

Questionnaire on Chemical Analysis of Soil

RESEARCH TRIANGLE INSTITUTE
POST OFFICE BOX 12194
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709

CHEMISTRY AND LIFE SCIENCES GROUP

October 28, 1980

Mr. William Mitchell
Toxicant Analysis Center
US Environmental Protection Agency
1105, NSTL
NSTL Station, Miss. 39529

Dear Mr. Mitchell:

The Research Triangle Institute (RTI), under contract with the Environmental Protection Agency (EPA), is conducting an assessment of the five National Pesticide Monitoring Programs. The Statistical Sciences Group at RTI has been analyzing the data generated by the Network Programs and is responsible for conducting this review. The Chemistry and Life Sciences Group is assuming a supportive role in this effort.

We have been asked to review the current analytical methodology being used and to evaluate the quality of data being generated in each Monitoring Program. The main objective of this review is not to criticize or find fault with the laboratories involved in these programs but to identify the strengths and limitations inherent in the analytical methodology. It is important to define the state-of-the-art as it is practiced by participating laboratories and to establish reliability factors for the reported data. The statisticians are particularly interested in assessing measurement error and in developing the best means for documenting estimates of accuracy.

We have prepared a list of questions relating to different aspects of the analytical procedure. Some questions are concerned with procedural matters and others are directed toward defining the scope of the methodology. We hope you will assist us by responding to these queries and by suggesting possible approaches or solutions to the issues mentioned above.

Since this evaluation must be based to some extent on your experience and view of the capabilities of the method, your cooperation is essential to the success of this evaluation. Your prompt response would be greatly appreciated.

Thank you.

Sincerely,

John W. Hines, Ph.D.
Chemist

JWH/lfo

(919) 541-6000 FROM RALEIGH, DURHAM AND CHAPEL HILL

National Soil Monitoring Program

-Analytical Methodology Issues-

1. It is presumed that current laboratory procedures follow a written analytical protocol. Please furnish a detailed copy of current laboratory protocol along with its source (e.g., EPA Manual of Analytical Methods). Include information on sample storage conditions (i.e., time, temperature) compositing) prior to analysis. Also include information on any procedural modifications required due to individual matrix or sample characteristics (e.g., emulsions, interferences or specific analytical requirements which might preclude the necessity for performing certain operations).

2. The following list represents compounds which have been monitored under the National Soil Monitoring Program. Please indicate which components are currently monitored on a routine basis, and which are no longer monitored or are only monitored under special circumstances (e.g., by request, in samples from particular geographic areas, in particular types of samples).

Organochlorines	Organophosphates	Heavy Metals
Alachlor	DEF	Mercury
Aldrin	Diazinon	Cadmium
BHC	Ethion	Lead
Chlordane	Malathion	Arsenic
DDTs	Phorate	
Dieldrin	Parathion, ethyl	
DCPA	Parathion, methyl	
Dicofol	Ronnel	
Endosulfan I	Trithion	
Endosulfan II		
Endosulfan Sulfate		
Endrin		
Endrin Ketone	Other	
Heptachlor	Trifluralin	
Heptachlor Epoxide		
Hexachlorobenzene		
Isodrin		
Lindane		
Methoxychlor		
PCBs		
Propachlor		
Toxaphene		

3. Which of the above analytes are never or very seldom found (<1% analyses) in general soil samples?

4. Do certain individuals perform specific aspects of the program (e.g., organophospate assays, data interpretation, QA/QC assessments)?

5. Are there "decision points" in your procedure where judgement is used in selecting procedural alternatives (e.g., column cleanup, choice of GC conditions, data interpretation)?

6. Describe your daily calibration and QC procedures (standards, spiked samples, blanks, other). Please indicate how many of each type of control sample are used with each sample set and their concentration levels (typical levels for standards, spiked samples).

7. Describe any additional QC/QA procedures which are part of your protocol (duplicate or split analysis, confirmatory analysis, use of multiple GC columns, interlaboratory programs, other). Please indicate how often these procedures are used.

8. What is the sample concentration range analyzed by direct injection on the GC, AA, etc (i.e., before further concentration or dilution becomes necessary)? Please indicate the method of reporting results at various analyte levels (i.e., above and below limit for quantitation, below limit of detection).

9. What are the estimates of the minimum quantitatable level (MQL) of individual analytes in real samples? How are they determined and to what extent does the sample matrix affect these values? What is the criterion used in reporting a specific analyte as "not detected" and in what manner are these results reported (zero, not detected, less than a certain value, less than the MQL)? Is the lower limit of quantitation different from the instrumental limit of detection? If so, what is their relationship?

10. What is your estimate of the analytical precision associated with each component and the dependence of this parameter on the analyte concentration in the sample? How is precision estimated? If available, please give the precision for analysis of replicate SPRMs or similar controls over a period of time for each analyte.

11. What is your estimate of the analytical accuracy associated with each component and the dependence of this paramter on the analyte concentration in the sample? How is accuracy estimated?

12. What is the analyte recovery during sample workup and is the reported concentration corrected for recovery?

13. What method(s) do you use for qualitative analysis of the data?

14. What method(s) do you use for quantitative analysis of the data?

15. What suggestions do you have for quantitating measurement error and documenting this information?

16. What suggestions do you have for making the Monitoring Program more efficient and meaningful (e.g., analytical modifications, choice of analytes for analysis, cost effectiveness)?

17. What are the number of person-hours (and costs, if possible) allocated for the sample workup, sample analysis, and data interpretation aspects of this program based on a set of samples? How many samples per set?

APPENDIX B

National Soil Monitoring Program
Pesticide Analysis Report Form

Table 1.3

SECTION 1. SAMPLE IDENTIFICATION DATA

LAB NO.	SAMPLE ACCESSION NO.	CD NO.	DATE RECEIVED AT LAB
1 2	3 4 5 6 7 8	9 10	
		0 7	

PESTICIDE ANALYSIS WORKSHEET

SECTION 2. SAMPLING DATA *(To be completed at sampling time)*

SAMPLED BY *(Agency and last name)*:

DATE SAMPLED	SITE			STATION/SITE NUMBER
11 12 13 14 15 16	STATE 17 18	COUNTY OR REGION 19 20 21		22 23 24 25 26 27 28 29 30 31

SYSTEM	MATERIAL	CROP NUMBER *(If applicable)*
32 33 NS = NATIONAL SOILS MONITORING NE = NATIONAL ESTUARINE MONITORING NW = NATIONAL WATER MONITORING	34 35 36	37

PESTICIDES USED *(Check or specify)*

2,4–D	CHLORDANE	DIELDRIN	MALATHION	TRIFLURALIN
ALDRIN	DDT	ENDRIN	PARATHION	
ATRAZINE	DIAZINON	HEPTACHLOR	TOXAPHENE	

SAMPLING REMARKS

SECTION 3. SPECIFIC SAMPLE CHARACTERISTICS

ESTUARINE — SPECIES *(Code)* 38 39 40 41 42 43

SOIL — pH 38 39 40 | % SAND 41 42 43 | % SILT 44 45 46 | % CLAY 47 48 49 | % ORGANIC MATTER 50 51 52

DATE ANALYSIS COMPLETED: 53 54 55 56 57 58

SECTION 4. RESIDUES DETECTED

9 10	PESTICIDE	CODE	AMOUNT	VAL*	9 10	PESTICIDE	CODE	AMOUNT	VAL*
1 8		11 12 13	14 15 16 17 18 19	20	0 9		11 12 13	14 15 16 17 18 19	20
		21 22 23	24 25 26 27 28 29	30			21 22 23	24 25 26 27 28 29	30
		31 32 33	34 35 36 37 38 39	40			31 32 33	34 35 36 37 38 39	40
		41 42 43	44 45 46 47 48 49	50			41 42 43	44 45 46 47 48 49	50
		51 52 53	54 55 56 57 58 59	60			51 52 53	54 55 56 57 58 59	60
		61 62 63	64 65 66 67 68 69	70			61 62 63	64 65 66 67 68 69	70
		71 72 73	74 75 76 77 78 79	80			71 72 73	74 75 76 77 78 79	80

M = P.P.M. *(default)*. B = P.P.B. *ug/kg whole body, wet weight*: T = P.P.T.

REMARKS

DATE							
ANALYST'S INITIALS							

EPA Form 8550-2 (Rev 2-75) PREVIOUS EDITIONS OBSOLET

APPENDIX C

Analytical Methodology for Organochlorine and Organophosphorous Pesticides and Trifluralin

Attached Methods

4.1 Extraction-Soil and Sediment

1. Weigh a 100 g specimen in a 500 ml Erlenmeyer flask and add 25 ml of distilled water.
2. Add 50 ml of nanograde acetone and place a teflon stopper in the flask. Shake specimen for ½ hour. Add 150 ml of nanograde hexane and continue shaking for 1½ hours more.
3. Decant specimen into a 500 ml separatory funnel through hexane-washed glasswool that has been baked at 350°C.
4. Wash the specimen 3 times with separate 100 ml portions of hexane-washed water. Discard the water (bottom layer) each time.
5. Pour the extract through a filter tube containing glasswool and a 1-inch layer of sodium sulfate that has been oven baked at 350°C. The filtrate is collected in a screw-capped test tube.
6. Store specimens in refrigerator until ready for use. The filtrate collected in step 5 is analyzed, without cleanup, for organophosphorus pesticides. Florisil cleanup is necessary for detection of organochlorine pesticides on the electron capture type of detector.
7. The moisture content of each specimen is determined by placing 100 g of soil sample in an oven at 125°C for 24 hours and then noting the weight loss of the sample.

Notes:

1. Run a solvent check with each group of specimens.
2. Run a fortified specimen with each group. The fortification procedure is as follows: Pipet 1.0 ml of the organochlorine "Soil Fortification Standard" A or 3.0l of a 1:3 dilution of "Soil

Fortification Standard" A into 100 g of soil or sediment specimen. Pipet 3 ml of the organophosphate "Soil Fortification Standard" into the same specimen. Mix the standards with the specimen and allow to stand overnight. The specimen is then extracted by the above procedure.

3. Dry weight = weight of specimen after heating overnight at 125°C.

A. Florisil Cleanup Procedure

1. Quantitatively transfer the specimen extract onto the top of the column and collect the elution from the column into a 250 ml flask.
2. When the sample extract drains down to the top of the upper layer of Na_2SO_4, add 100 ml of a mixture consisting of 10% methylene chloride in hexane and continue collecting until the liquid level reaches the upper Na_2SO_4 layer. This elution is labeled the "first fraction."
3. Replace the first 250 ml flask with a second flask and then add 100 ml of 100% nanograde methylene chloride to the Florisil column. Continue collecting the elution until the column drains dry. Label the eluted portion, "fraction two."
4. To each flask add 1.0 ml of 0.01% Nujol (in hexane) and 3 to 4 glass beads. Attach a 3-ball Snyder column and place on a steam bath or hotplate. Concentrate to ca 5 ml. Add 50 ml nanograde hexane and concentrate to about 5 ml. Repeat the last concentration step once more. This will remove essentially all methylene chloride.
5. Pour 5 ml of hexane through the top of the Snyder column (for rinsing) and collect in the flask.
6. Transfer specimens quantitatively into 15 ml graduated centrifuge tubes and place into a water bath that is maintained at 40°C.
7. Direct a purge of air into the centrifuge tube above the liquid level until the volume of liquid is reduced to 2.5 ml.
8. Samples are now ready for CG determination.

F. Concentration of Specimens on Hot Plate

1. Swirl the flasks containing glass beads until boiling occurs.
2. Do not allow the flasks to evaporate to dryness.

G. Pouring of Extracts Into Graduated Centrifuge Tubes

1. Use a small funnel to avoid losses due to direct pouring.

H. Concentration of Samples In Centrifuge Tubes With A Stream of Dry Air

1. Water bath should remain at a constant temperature.
2. Stream of air to all samples should be about the same flow rate.
3. Concentrate all samples to approximately the same volume.

I. Column Cleanup

1. It is important that the adsorbent (Florisil) have consistent mesh size and moisture content.
2. Exactly the same weight of adsorbent should be used for each sample.
3. Good column technique is essential for adequate separations.

C. Florisil Column Separation of Pesticides in Standards A and B

1. Components eluting in the first fraction (150 ml of 10% methylene chloride in hexane) are:

    aldrin

    heptachlor

    gamma chlordane

    OPDDE

    PPDDE

    OPDDT

    PPDDT

    PPTDE

2. Components eluting in the second fraction (100 ml of methylene chloride) are:

    endrin

    dieldrin

    *heptachlor epoxide

*occasionally heptachlor epoxide may split between the two fractions.

D. Florisil Column Separation of Other Common Pesticides

1. First fraction

trifluralin

toxaphene

PCB's

lindane (BHC)

PCNB

chlordane

methoxychlor

mirex

2. Second fraction

endosulfan I

endosulfan II

endosulfan sulfate

endrin, aldehyde form

endrin, ketone form

Note:

Most organophosphorus pesticides elute in the second fraction.

B. Each Batch of Florisil Should Be Checked As Follows:

1. Add known volume of bench standard to Florisil column, and take off fractions, as in the above procedure.
2. Concentrate volumes of fractions 1 and 2 to the same volume as that originally added to the Florisil column.
3. Compare recoveries in each fraction with the bench standard. This allows the chromatographer to determine which fraction contains each component and the percent loss on the Florisil column, if any.

APPENDIX D

Sampling Weights for the Rural Soils Network (RSN)

0. NOTATION

1967 Conservation Needs Inventory (CNI)

National Soil Monitoring Program (NSMP)

Rural Soils Network (RSN)

Rural Soils Network Cropland Sample (RSN_1)

Rural Soils Network Noncropland Sample (RSN_2)

Let $i = 1, \ldots, 48$	denote the States of the conterminous United States
Let $j = 1, \ldots, s(i)$	denote the counties of State i that are not strictly metropolitan in character
Let k[1/] $= 1, \ldots, t(i,j)$	denote the strata in county j of State i
Let $\ell = 1, \ldots, U(i,j,k)$	denote the primary sampling units (PSU's), typically 160- acre plots, in stratum k of county j in State i
Let $\ell = 1, \ldots, u(i,j,k)$	denote the sample PSU's in stratum k of county j in State i

[There are uncountably many secondary sampling units (SSU's), i.e. possible sampling points, in each PSU, so it is not possible to index the population of SSU's within any PSU.]

Let $m = 1, \ldots, v(i,j,k,\ell)$	denote the actual SSU's selected by spinning the sampling template once for PSU ℓ in stratum k of county j in State i.

[1/] Although townships or their equivalent are used to stratify the sample within counties, the township, within township, and other levels of stratification are treated herein as a single level without loss of generality.

Let $V(i,j,k,\ell)$ be the random variable representing the number of SSU's selected by spinning the sampling template for PSU ℓ in stratum k. Note that $v(i,j,k,\ell)$ is a realization of $V(i,j,k,\ell)$.

Let $m_1 = 1, \ldots, v_1(i,j,k,\ell)$ denote the realized cropland SSU's for PSU ℓ in stratum k of county j in state i

Let $m_2 = 1, \ldots, v_2(i,j,k,\ell)$ denote the realized noncropland SSU's for PSU ℓ in stratum k of county j in State i.

Of course, $v_1(i,j,k,\ell) + v_2(i,j,k,\ell) = v(i,j,k,\ell)$.

1. PHASE ONE -- THE CNI SAMPLE

1.1 CNI PSU Probability

Since $u(i,j,k)$ PSU's are selected at random and without replacement from the $U(i,j,k)$ PSU's in stratum (i,j,k),

$p(i,j,k)$ = Overall probability of selection into the CNI for each PSU in stratum (i,j,k)

$$= \frac{u(i,j,k)}{U(i,j,k)} . \qquad (1)$$

For the standard sampling procedure, in which one PSU was selected at random from a stratum containing 48 PSU's,

$$p(i,j,k) = \frac{u(i,j,k)}{U(i,j,k)} = \frac{1}{48} \doteq 2\% .$$

1.2 Conditional Probability for SSU's in the CNI

Recall that $m = 1, \ldots, v(i,j,k,\ell)$ indexes the CNI sample points in PSU ℓ of stratum k. Also recall that there are infinitely many such points available for sampling in each PSU. If the points are

considered to have no dimensions and hence no area, any point picked at random must have zero probability of being selected into the CNI sample. This is because there are infinitely many mutually exclusive points.

However, a point with no dimensions cannot be assigned a land use other than that of a small undefined physical area surrounding that point. Thus, in fact, a small undefined area centered at each CNI sampling point was sampled rather than a point, per se. Let us then assume that each CNI sample point is effectively a sampling unit with area a, where the area a does not depend upon PSU or stratum. A probability density for sample selection can then be distributed over each PSU, resulting a positive probability for each SSU.

A reasonable simplification seems to be to assume that the probability density for selection is uniform over each PSU. This assumption would imply, among other things, that there is no border effect. That is, areas near the edge of the PSU are neither over- nor under-represented in the sample, both as selected and as implemented in the field. In this case, if a single SSU were to be selected at random within a sampled PSU, its conditional probability of selection would be a/A, where A is the area of the PSU and a is the area of the SSU.

A random number V (i,j,k,ℓ) of SSU's were selected from PSU(i,j,k,ℓ). Letting A(i,j,k,ℓ) denote the area of PSU(i,j,k,ℓ), the conditional probability, given selection of PSU(i,j,k,ℓ), for the selection of SSU (i,j,k,ℓ,m) is then

$$\text{Prob [SSU } (i,j,k,\ell,m) \text{ is selected / PSU } (i,j,k,\ell) \text{ is selected]}$$

$$= \frac{a}{A(i,j,k,\ell)} \, E\,[V(i,j,k,\ell) \text{ / PSU } (i,j,k,\ell) \text{ is selected}]. \quad (2)$$

The expected number of sample SSU's in a PSU, i.e. E[V] in (2), is proportional to the area, A(i,j,k,ℓ), of the PSU (except for 640 acre PSU's). The density of the sampling template for 640-acre PSU's, adjusted to a common photograph scale, was one-fourth that for all other PSU's. Hence, the proportionality constant for 640-acre PSU's is one-fourth of that for all other PSU's, hence,

$$E[V(i,j,k,\ell) \mid \text{PSU } (i,j,k,\ell) \text{ is selected}] = \begin{cases} 0.25\ c\ A(i,j,k,\ell) & \text{for 640-acre PSU's} \\ 1.00\ c\ A(i,j,k,\ell) & \text{for all other PSU's} \end{cases} \quad (3)$$

Thus, from (2) and (3)

$$\text{Prob [SSU } (i,j,k,\ell,m) \text{ is selected} \mid \text{PSU } (i,j,k,\ell) \text{ is selected}] = \begin{cases} 0.25c\ a & \text{for 640-acre PSU's} \\ 1.00c\ a & \text{for all other PSU's} \end{cases} \quad (4)$$

That is, the conditional probability of selection of an SSU is a constant that depends only upon size of the PSU.

1.3 CNI Sampling Weights

Combining the results of 1.1 and 1.2, we can determine the overall probability of selection for the ultimate sampling units, the SSU's, for the CNI sample. In particular, it follows from (1) and (4) that

$$\text{Prob [SSU } (i,j,k,\ell,m) \text{ is selected into the CNI sample}] = \text{Prob [PSU } (i,j,k,\ell) \text{ is selected into the CNI sample}] \times \text{Prob [SSU } (i,j,k,\ell,m) \text{ is selected} \mid \text{PSU } (i,j,k,\ell) \text{ is selected}]$$

$$= \begin{cases} 0.25\ a\ c\ p(i,j,k) & \text{for 640-acre PSU's} \\ 1.00\ a\ c\ p(i,j,k) & \text{for all other PSU's} \end{cases} \quad (5)$$

Thus, for estimation of means, a proper sampling weight for each SSU record in the CNI sample is simply

$$W(i,j,k,\ell,m) = \begin{cases} \dfrac{4}{p(i,j,k)} & \text{for 640-acre PSU's} \\ \dfrac{1}{p(i,j,k)} & \text{for all other PSU's} \end{cases} \qquad (6)$$

The constant factor, ac, cancels in any estimation of means. Of course, this weight reflects only the unequal probabilities of selection due to the sampling design and can be further modified to reflect missing data, failure to accurately locate sampling points, etc.

1.4 Weighing the CNI to Estimate Total Land Area

It seems reasonable that if each SSU of the CNI is to be regarded as having area equal to one (unit free), then the sampling weight to be assigned to an SSU is

$$WT(i,j,k,\ell,m) = \frac{E[\text{Area (in acres) represented by the SSU}]}{\text{Prob [PSU } (i,j,k,\ell)]}$$

$$= \frac{\dfrac{\text{Area of PSU } (i,j,k,\ell)}{E[V(i,j,k,\ell)]}}{\text{Prob [PSU } (i,j,k,\ell)]} \cdot$$

$$= \begin{cases} \dfrac{\dfrac{A(i,j,k,\ell)}{0.25\ c\ A(i,j,k,\ell)}}{p(i,j,k)} \cdot & \text{for 640-acre PSU's} \\ \dfrac{\dfrac{A(i,j,k,\ell)}{c\ A(i,j,k,\ell)}}{p(i,j,k)} \cdot & \text{for all other PSU's} \end{cases}$$

$$= \begin{cases} \dfrac{4}{c\ p(i,j,k)} & \text{for 640-acre PSU's} \\ \dfrac{4}{c\ p(i,j,k)} & \text{for all other PSU's} \end{cases} \qquad (7)$$

Of course, the proportionality constant, c, or equivalently, $E[V(i,j,k,\ell)]$ would have to be explicitly determined, probably empirically, to actually use (7) in estimation of total acreage. Although we will not need (7) explicitly, since we are only interested in estimating means or rates, it is reassuring that the weights (6) and (7) are of the same form.

2. THE RSN SAMPLE

2.1 Preliminaries for the RSN

The contribution of sampled CNI cropland PSU (i,j,k,ℓ) to the cropland accumulation used by the USDA for selecting the RSN subsample is the adjusted cropland ratio

$$r_1(i,j,k,\ell) = \frac{v_1(i,j,k,\ell)}{v(i,j,k,\ell)} \cdot \frac{0.02}{p(i,j,k)} . \tag{8}$$

Thus, the total of the cropland accumulation used by the USDA in State i is

$$N_1(i) = \sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} \sum_{\ell=1}^{u(i,j,k)} \frac{v_1(i,j,k,\ell)}{v(i,j,k,\ell)} \frac{0.02}{p(i,j,k)}$$

$$= 0.02 \sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} \frac{1}{p(i,j,k)} \sum_{\ell=1}^{u(i,j,k)} \frac{v_1(i,j,k,\ell)}{v(i,j,k,\ell)} \tag{9}$$

Similarly, the total of the noncropland accumulation in State i is

$$N_2(i) = 0.02 \sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} \frac{1}{p(i,j,k)} \sum_{\ell=1}^{u(i,j,k)} \frac{v_2(i,j,k,\ell)}{v(i,j,k,\ell)} \tag{10}$$

2.2 Estimation of Proportion of Cropland Acreage in the Rural Area of State i.

The estimate used was

$$\frac{N_1(i)}{N_1(i) + N_2(i)} = \frac{\sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} \frac{1}{p(i,j,k)} \sum_{\ell=1}^{u(i,j,k)} \frac{v_1(i,j,k,\ell)}{v(i,j,k,\ell)}}{\sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} \frac{1}{p(i,j,k)} u(i,j,k)}$$

$$= \frac{\sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} \frac{1}{p(i,j,k)} \sum_{\ell=1}^{u(i,j,k)} \frac{v_1(i,j,k,\ell)}{v(i,j,k,\ell)}}{\sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} U(i,j,k)} \tag{11}$$

= (Estimated total of the cropland proportions for all PSU's in State i) ÷ (Total number of PSU's in State i)

2.3 More Preliminaries for the RSN

Let $n_1(i)$ denote the number of 10-acre RSN_1 sites to be selected in State i. Recall that $n_1(i)$ is chosen so that 0.025% of the cropland acreage in State i is sampled.

The procedure for selecting $n_1(i)$ starting points for the $n_1(i)$ RSN_1 sample sites was:

1) Select a random number from the interval $(0, w_1(i))$ where $w_1(i) = N_1(i)/n_1(i)$. Call this random number $q_1(i)$.
2) Select as an RSN_1 starting point the first CNI cropland SSU whose contribution to the cropland accumulation causes the accumulation to equal or exceed $q_1(i)$.
3) Repeat step (2) with $q_1(i)$ replaced by $q_1(i) + w_1(i)$, $q_1(i) + 2\,w_1(i)$, . . ., $q_1(i) + [n_1(i)-1]\,w_1(i)$.

It should be noted that an RSN_1 starting point did not uniquely determine an RSN_1 sample site. In particular, an adjacent cropland SSU had to be found, and the RSN_1 sample site was centered about these two cropland SSU's from the CNI. Moreover, substitution procedures were employed when an adjacent cropland SSU did not exist. In addition, a substitute RSN_1 site was selected if the selected site either could not be surveyed for some reason or could no longer be considered a cropland site.

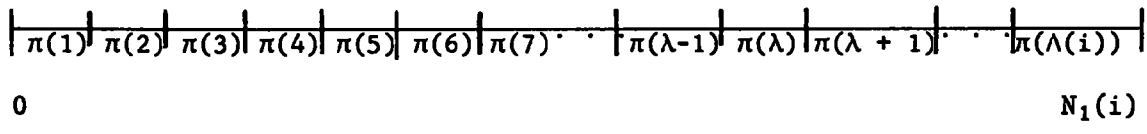
Let us consider an alternate, and perhaps more useful, representation of identically the same procedure for selecting the $n_1(i)$ starting points for the RSN_1 sample. The contribution of each cropland SSU in PSU (i,j,k,ℓ) to the cropland accumulation is

$$\frac{1}{v\ (i,j,k,\ell)} \qquad \frac{0.01}{p\ (i,j,k)} \tag{12}$$

Let $\lambda = 1, \ldots, \Lambda(i)$ denote the SSU's selected into the CNI sample in State i. The cropland accumulation may then be represented as

$$N_1(i) = \sum_{\lambda=1}^{\Lambda(i)} \pi(\lambda) \quad , \tag{13}$$

where $\pi(\lambda)$ is given by (12) for cropland SSU's and is zero for noncropland SSU's. Thus, the cropland accumulation for State i may be thought of as partitioned into $\Lambda(i)$ zones, where each zone has width $\pi(\lambda)$. PSU's that are entirely noncropland will contribute a null zone with zero width to the cropland accumulation. The RSN_1 procedure for selecting a CNI SSU as a starting point for a 10-acre RSN_1 site may then be illustrated as:



A cropland SSU is selected as a starting point for locating an RSN_1 site, if the sequence number

$$q_1(i) + \gamma w_1(i) \quad \text{for } \gamma = 0, 1, \ldots, n_1(i) - 1 \tag{14}$$

hits the zone representing the SSU.

2.4 Conditional probabilities for SSU's in the RSN_1 given the CNI

The $SSU(i,j,k,\ell,m_1)$ is selected as an RSN_1 starting point only if the single random number $q_1(i)$ results in a sequence number given by (14) that hits the zone representing $SSU(i,j,k,\ell,m_1)$. The chance of multiple hits on this zone is almost identically zero since the width of the zone representing a cropland SSU, given by (12), is very much smaller than the distance $w_1(i) = N_1(i)/n_1(i)$ between cropland sequence numbers.[2/] Thus,

2/ Multiple hits within the same PSU have occurred in the RSN sample occasionally, however, due to the inadvertent repetition of some PSU records in the State lists used to select the RSN subsamples.

Prob (SSU (i,j,k,ℓ,m_1) will be selected as a starting point for an RSN_1 site SSU (i,j,k,ℓ,m_1) is in the CNI sample)

$$= \frac{\text{Size of the zone representing SSU } (i,j,k,l,m_1)}{w_1(i)}$$

$$= \frac{\frac{1}{v\ (i,j,k,\ell)} \quad \frac{0.02}{p\ (i,j,k)}}{N_1(i)/n_1(i)} \tag{15}$$

from (12) and the fact that $q_1(i)$ is a random number from the interval $(0,\ w_1(i))$.

2.5 RSN sampling weights

It will be recalled that the selection of a starting point for locating a 10-acre RSN_1 site did not uniquely specify the site. There was a procedure for determining the sampling site based on any cropland starting point, however, as long as an appropriate site could be found within the PSU containing the starting point. To the extent that this procedure was strictly applied, most RSN_1 sites were uniquely determined. However, it is apparent from considering several maps of RSN_1 sites that the specified procedure was only adhered to loosely.

It should also be noted that the procedure for determining an RSN_1 site based upon a cropland starting point was that the starting point not be included in the resulting sample site if the starting point was an isolated cropland point. Thus, there was an intentional bias away from isolated cropland SSU's in the RSN_1.

If the non-uniqueness of the RSN_1 site determined by selection of a starting point, and bias away from isolated cropland SSU's is ignored, we obtain from (5) and (15),

Prob (the RSN_1 site resulting from selection of SSU (i,j,k,ℓ,m_1) as a starting point will be selected into the RSN_1 sample)

= Prob (SSU(i,j,k,ℓ,m_1) will be selected as a starting point for an RSN_1 site / SSU(i,j,k,ℓ,m_1) is in the CNI sample)

× Prob (SSU (i,j,k,ℓ,m_1) is selected into the CNI sample)

$$= \begin{cases} \dfrac{1}{\dfrac{v\ (i,j,k,\ell)}{N_1(i)/n_1(i)}} \times \dfrac{0.02}{p(i,j,k)} & \begin{array}{l} 0.25\ a\ c\ p\ (i,j,k) \text{ for} \\ 640\text{-acre PSU's} \\ a\ c\ p(i,j,k) \text{ for} \\ \text{all other PSU's} \end{array} \\ \dfrac{0.005\ a\ c\ n_1\ (i)}{v(i,j,k,\ell)\ \ N_1(i)} & \text{for 640-acre PSU's} \end{cases}$$

$$= \qquad (16)$$

$$= \frac{0.02\ a\ c\ n_1(i)}{v\ (i,j,k,l)\ \ N_1(i)} \qquad \text{for all other PSU's}$$

Thus, if we are willing to accept the simplying assumptions at the beginning of this section, a sampling weight for estimation of means for the RSN_1 site resulting from selection of SSU (i,j,k,ℓ,m_1) as a starting point is

$$W_1\ (i,j,k,\ell,m_1) = \begin{cases} \dfrac{4v\ (i,j,k,\ell)\ \ N_1(i)}{n_1(i)} & \text{for 640-acre PSU's} \\ \dfrac{v\ (i,j,k,\ell)\ \ N_1(i)}{n_1(i)} & \text{for all other PSU's.} \end{cases} \qquad (17)$$

It should be noted that the weight given by (17) will be approximately the same for all RSN_1 sites within those States where only one PSU size was used. This is because $v(i,j,k,\ell)$ will be very nearly the same for all PSU's within such a State.

Of course, a sampling weight for estimation of means for the RSN_2 site resulting from selection of SSU(i,j,k,ℓ,m_2) as a starting point is

$$W_2\ (i,j,k,\ell,m_2) = \begin{cases} \dfrac{4\ v(i,j,k,\ell)\ N_2(i)}{n_2(i)} & \text{for 640-acre PSU's} \\ \dfrac{v(i,j,k,\ell)\ N_2(i)}{n_2(i)} & \text{for all other PSU's.} \end{cases} \qquad (18)$$

3. Comments

A strict accounting of the bias away from isolated cropland SSU's in the RSN_1 would be quite difficult. It would be necessary to determine, for each RSN_1 site, the number of SSU's that would have resulted in selection of the site if that SSU had been chosen as a starting point. This number of SSU's chosen as starting points that would have resulted in selection of the RSN_1 site could theoretically be any positive integer. A value of one would hopefully be predominant, giving exact aggrement with (17) and (18). However, two and three would surely occur also.

Consider the conditional probabilities for the SSU's in the RSN_1, given the CNI, as considered in Section 2.4. In particular, the sum over all SSU's sampled by the CNI of the probability that $SSU(i,j,k,\ell,m_1)$ will be selected as a starting point for RSN_1 site is as follows from (15):

$$\sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} \sum_{\ell=1}^{u(i,j,k)} \sum_{m_1=1}^{v_1(i,j,k,\ell)}{}^{*} \frac{1}{N_1(i)/n_1(i)} \frac{1}{v(i,j,k,\ell)} \frac{0.02}{p(i,j,k)}$$

$$= \sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} \sum_{\ell=1}^{u(i,j,k)} \frac{n_1(i)}{N_1(i)} \frac{0.02}{p(i,j,k)} \frac{v_1(i,j,k,\ell)}{v(i,j,k,\ell)}$$

$$= \frac{0.02\, n_1(i)}{N_1(i)} \sum_{j=1}^{s(i)} \sum_{k=1}^{t(i,j)} \frac{1}{p(i,j,k)} \sum_{\ell=1}^{u(i,j,k)} \frac{v_1(i,j,k,\ell)}{v(i,j,k,\ell)}$$

$$= n_1(i)$$

from (9). Thus, the sum of the SSU probabilities for the RSN_1, conditional on the CNI sample being regarded as fixed, is the RSN_1 sample size for State i, namely $n_1(i)$. This result lends additional credence to the correctness of the sampling weights as described by (17).

*Summation is over the sample cropland points of the CNI sample because these points constitute the population with regard to the conditional RSN probabilities.

4. Approximation to the RSN Sampling Weights

Exact implementation of the sampling weights given by (17) and (18) is not a simple task. The sample sizes $n_1(i)$ and $n_2(i)$ for the cropland and noncropland samples are readily available (See Table 1.3). However, the State accumulations of the adjusted cropland and noncropland ratios, $N_1(i)$ and $N_2(i)$, are only available from the hard-copy computer records of the RSN sample selection. These records are not entirely reliable, since there is no guarantee that the copy available was the final copy from which the sample was selected. Dummy records were added to obtain coverage of federal croplands for the noncropland sample, and the data set was otherwise edited before sample selection. The number of sampling points, $v(i,j,k,\ell)$, is again available from the hard-copy computer records. However, it would be a monumental task to go through the hard-copy computer records to obtain $v(i,j,k,\ell)$ for each RSN sampling site. A perusal of these sampling records reveals that individual CNI sites were sometimes entered more than once, doubling the probability that these sites would enter the RSN sample.

If the sampling design had been implemented exactly as described in the text for a particular State and all PSU's were 160 acres, partial PSU's would still occur around the boundaries of counties and other large scale geographic strata, e.g., irrigated areas. These partial PSU's would be "nominal" 160-acre PSU's, but would receive fewer than the usual number of sampling points.

The full 160-acre PSU's each receive approximately 36 sampling points. The random variation in $v(i,j,k,\ell)$ may be small for the full PSU's, and a good approximation to the sampling weights given by (17) and (18) is achieved by using the mean number of points assigned in place of $v(i,j,k,\ell)$.

Identification of the "nominal" 160-acre PSU's is not be a simple task. It requires close examination of the CNI sampling maps, at least. It should be noted also that actual PSU's were, in practice, sometimes larger than their "nominal" size. These larger PSU's occurred mostly in States that used 40-acre PSU's in "irrigated" strata, where the "nominal" 40-acre PSU's were sometimes larger than 40 acres around the stratum boundaries. Due to the problem of identifying PSU's considerably larger or smaller than their "nominal" size, no adjustment in the sampling weights (17) and (18) is being proposed for RSN sites occuring in these PSU's.

The sampling weights given by (17) and (18) are only appropriate if the sampling design is implemented as described in the text. Examination of the numbers of points assigned to CNI sites reveals, however, that this was not the case. The assignment of sampling points within PSU's was done at local USDA offices, and the design sampling protocol was not consistently followed. For example, nearly all sites in Nevada received approximately 36 sampling points, whether the PSU size was 40 acres or 160 acres. Moreover, it appears that the sampling template may not have been spun for Nevada sites since most received exactly 36 sampling points. Also, the scales of the sampling template and the aerial photograph were often not properly matched, resulting in consistently more or fewer sampling points than expected from the design protocol. For example, many 160-acre PSU's in New Mexico received approximately 18 sampling points, rather than 36 sampling points. Thus, a single sampling protocol was not consistently applied throughout the United States. It is probably not possible to determine exactly what protocol was used for each sampling site. The consequences of these variations in sampling protocol will presently be investigated.

The investigation of the effects of variations in the CNI sampling procedure upon the RSN sampling weights will be aided by considering the sampling weight given by (17) as a product. In particular, the sampling weight (17) may be written as

$$W_1(i,j,k,\ell,m_1) = \begin{cases} \dfrac{v(i,j,k,\ell)\ p(i,j,k)\ N_1(i)}{n_1(i)} \cdot \dfrac{4}{p(i,j,k)} & \text{for 640-acre PSU's} \\ \dfrac{v(i,j,k,\ell)\ p(i,j,k)\ N_1(i)}{n_1(i)} \cdot \dfrac{1}{p(i,j,k)} & \text{for all other PSU's} \end{cases}$$

$$= \begin{cases} \dfrac{v(i,j,k,\ell)\ N_1(i)}{n_1(i)} \cdot 4 & \text{for 640-acre PSU's} \\ \dfrac{v(i,j,k,\ell)\ N_1(i)}{n_1(i)} \cdot 1 & \text{for all other PSU's,} \end{cases} \qquad (19)$$

where the first factor is the conditional RSN weight and the second factor is the CNI weighting factor. The CNI weighting factor for 640-acre PSU's is four times that for all other PSU's because each such point represents four times as much land area as points in PSU's of other sizes.

A specific case may help to clarify the effects of variations in the CNI sampling procedure upon the RSN sampling weights. Once again, consider the case of New Mexico where many 160-acre PSU's were sampled at the design rate of about 36 sampling points per PSU, while many other 160-acre PSU's were sampled at the lower rate of about 18 points per PSU. For those PSU's sampled at the proper intensity, the appropriate sampling weight is approximately

$$\frac{36\ N_1(i)}{n_1(i)} \cdot 1 \quad .$$

The sampling weight formula given by (19) would have to be modified for the PSU's receiving only about 18 sampling points, since each point represents twice as much land area. The CNI factor of the sampling weight is doubled resulting in an approximate RSN sampling weight of

$$\frac{18\ N_1(i)}{n_1(i)} \cdot 2 = \frac{36\ N_1(i)}{n_1(i)} ,$$

which is exactly the same as the first case. When half the usual number of sampling points was assigned the conditional RSN weight was halved. However, each sampling point then represented twice as much land area, doubling the CNI weighting factor. In terms of probability, the conditional RSN probability was doubled for each point, since it contributed twice as much to the accumulation $N_1(i)$, but the unconditional CNI probability was halved, since half as many sampling points were being assigned within the PSU. Thus, the procedural variations in the CNI sampling protocol result in no change in the appropriate mean weight for the RSN. A single mean sampling weight is then appropriate for all 160-acre PSU's. This sampling weight is

$$\frac{\bar{v}_{160}\ N_1(i)}{n_1(i)} ,$$

where $\bar{v}_{160}$ is the average number of sampling points per PSU when the design sampling procedure described in the text is applied.

This weight fails to reflect the random variation in the number of sampling points, v, assigned to a PSU within any given sampling protocol. However, it is a proper mean sampling weight regardless of the sampling protocol. The alternative is not feasible, requiring precise knowledge of the sampling procedure used to assign the sampling points as well as the number of points assigned for each PSU containing an RSN sample site. Since this detailed information is not available, the mean sampling weights appear to be most appropriate.

The following mean weights are suggested for the RSN cropland sample:

$$\text{40 acre PSU's} : \frac{\bar{v}_{40}\, N_1(i)}{n_1(i)}$$

$$\text{100 acre PSU's} : \frac{\bar{v}_{100}\, N_1(i)}{n_1(i)}$$

$$\text{160 acre PSU's} : \frac{\bar{v}_{160}\, N_1(i)}{n_1(i)}$$

$$\text{640 acre PSU's} : \frac{\bar{v}_{640}\, N_1(i)}{n_1(i)} \quad ,$$

where $\bar{v}_A$ is the mean number of sampling points assigned to PSU's of area A under the sampling protocol specified by the design. It should be noted, however, that this protocol results in $\bar{v}_A$ being directly proportional to the size, A, of the PSU, except for 640-acre PSU's where $\bar{v}_{640}$ is identical to $\bar{v}_{160}$. Thus, the above mean RSN sampling weights may be expressed as follows:

$$\text{40-acre PSU's} : \frac{4\, \bar{v}_{10}\, N_1(i)}{n_1(i)}$$

$$\text{100-acre PSU's} : \frac{10\, \bar{v}_{10}\, N_1(i)}{n_1(i)}$$

$$\text{160-acre PSU's} : \frac{16\, \bar{v}_{10}\, N_1(i)}{n_1(i)}$$

$$\text{640-acre PSU's} : \frac{4\, \{16\, \bar{v}_{10}\}\, N_1(i)}{n_1(i)} = \frac{64\, N_1(i)}{n_1(i)} \quad ,$$

where $\bar{v}_{10}$ is the mean number of sampling points per 10 acres assigned to all but 640-acre PSU's under the sampling protocol specified by the design.

Since only relative sampling weights are required for estimation of means, the constant factor, $\bar{v}_{10}$, in the above sampling weights may be

cancelled. Moreover, the cropland and noncropland samples of the RSN can be regarded as two strata in the RSN sample of the rural areas of the conterminous United States. As seen before, the derivation of the noncropland sampling weights parallels that for the cropland sampling weights in all respects. The ratio $N_1(i)/n_1(i)$ for the cropland sampling weights in state i is replaced by $N_2(i)/n_2(i)$ for the noncropland sample. Otherwise, the conditional RSN factor of the sampling weights and the unconditional CNI factor remain unchanged. Thus, the final recommended sampling weights for the RSN are as given in Table D-1. The cropland sampling rate was 0.025 percent of the cropland acreage within each state, and the noncropland sampling rate was 0.0025 percent of the noncropland acreage, which is reflected by $N_2(i)/n_2(i)$ being approximately 10 times as large as $N_1(i)/n_1(i)$.

Table D-1: Recommended RSN Sampling Weights

PSU Size	Cropland	Noncropland
40 acres	$4\ N_1(i)/n_1(i)$	$4\ N_2(i)/n_2(i)$
100 acres	$10\ N_1(i)/n_1(i)$	$10\ N_2(i)/n_2(i)$
160 acres	$16\ N_1(i)/n_1(i)$	$16\ N_2(i)/n_2(i)$
640 acres	$64\ N_1(i)/n_1(i)$	$64\ N_2(i)/n_2(i)$

Notation: $n_1(i)$ = Number of cropland sample sites in state i

$n_2(i)$ = Number of noncropland sample sites in state i

$N_1(i)$ = Total "cropland accumulation" for state i

$N_2(i)$ = Total "noncropland accumulation" for state i

It should be emphasized that the sampling weights shown in Table D-1 reflect only the mean differences in the portion of the selection probabilities of the RSN sites that depend upon the size of the PSU. It has been argued that the selection probabilities would be fairly constant for a given size of PSU, since the total number of CNI sampling points would be fairly constant. Undocumented variations in the CNI sampling

protocol make it virtually impossible to quantify the smaller variations in selection probabilities for RSN sample sites within the group of PSU's of a given size. There are many other factors that may be reflected in sampling weights, but are presently ignored. Some of these factors are:

1) Duplicate entry of some CNI sites in the list from which the RSN sample was selected, doubling the chance of selection for all potential RSN sites within such PSU's.

2) CNI sites missed when the RSN sample was selected.

3) Inclusion of some CNI sites that fell outside the partial PSU being sampled.

4) Loss of some CNI site maps.

5) PSU's substantially over or under their "nominal" size.

6) Border effects, or PSU size effects, on the number of sampling points assigned within a PSU.

7) Random variation in the total number of CNI sampling points assigned to a PSU.

8) Failure to accurately locate the selected RSN sites, CNI sites, and/or CNI sampling points in the field.

9) Uncertainty associated with the values found for the cropland and the noncropland accumulations for each state.

10) The existence of multiple CNI sampling points that would all lead to selection of the same RSN site.

11) The use of substitute RSN sites.

5. Implementation of the RSN Sampling Weights

Implementation of the approximate RSN sampling weights shown in Table D-1 required that information be gathered that was not available on the data records. The size of the PSU from the 1967 CNI sample into which each RSN site fell was obtained from the Statistical Laboratory at Iowa State University. These findings are shown in Tables D-2 through D-5. The State accumulations of the adjusted cropland and noncropland

Table D-2: States With Only 160-Acre PSU's*

State Code	State Name	State Code	State Name
01	Alabama	28	Mississippi
06	California	29	Missouri
08	Colorado	30	Montana
12	Florida	37	North Carolina
13	Georgia	38	North Dakota
16	Idaho	39	Ohio
17	Illinois	40	Oklahoma
18	Indiana	41	Oregon
19	Iowa	45	South Carolina
20	Kansas	47	Tennessee
21	Kentucky	48	Texas
22	Louisiana	53	Washington
26	Michigan	55	Wisconsin
27	Minnesota		

* Source: Statistical Laboratory, Iowa State University

Table D-3: States With Only 100-Acre PSU's*

State Code	State Name	State Code	State Name
09	Connecticut	36	New York
10	Delaware	42	Pennsylvania
24	Maryland	44	Rhode Island
25	Massachusetts	50	Vermont
33	New Hampshire	51	Virginia
34	New Jersey	54	West Virginia

* Source: Statistical Laboratory, Iowa State University

Table D-4: States With Constant PSU Size Within Counties*

State Code	State Name	PSU Size	County Codes
05	Arkansas	160 acres	9-15,23,39,49,53,59,61,89, 99,103,109,129,133-137
		40 acres	All others
46	South Dakota	640 acres	7,19,31,33,41,47,55,63,71, 75,81,85,93,95,103,105,113, 117,121,131,137
		160 acres	All others

* Source: Statistical Laboratory, Iowa State University

Table D-5: States With Varying PSU Size Within Counties*

State Code	State Name	PSU Size	RSN Site Numbers[1/]
23	Maine	400 acres	7,27,29,32-34,36-39,48,61,67
		100 acres	All others
04	Arizona	160 acres	1,3,10-55,107,108,160,163
		40 acres	All others
31	Nebraska	640 acres	67,68,70,180,194,195,243, 246-248,321-324,434,448,449, and all sites in counties: 5,9,31,41,45,63,73,75,91, 117,161,165,171
		160 acres	All others
32	Nevada	160 acres	96,143
		40 acres	All others
35	New Mexico	640 acres	65,67,179
		40 acres	1,4,59,62,64,117,120,123,125, 177,183,184
		160 acres	All others
49	Utah	640 acres	12,51,91,97
		160 acres	2,3,9,45,46,48,52,54,56,90, 92,95,134,135,139,141
		40 acres	All others
56	Wyoming	640 acres	4-6,9,15,17-20,22-25,30-33, 35-43,46,48-54,57,60,61,66, 71,112,165,167,168,174,176
		160 acres	All others

1/ Only sites for which data was collected were classified. Completion of this table for all RSN sample sites in these states would be very time consuming.

* Source: CNI site numbers corresponding to the RSN site numbers were obtained from the EPA Field Studies Branch, Washington, D.C. The PSU size for each of these CNI sites was obtained from the Statistical Laboratory at Iowa State University.

ratios, $N_1(i)$ and $N_2(i)$, were obtained from the hard-copy computer records of the RSN sample selection. The information obtained is shown in Table D-6. The information in Tables D-2 through D-6 was then used for the sampling weight computations shown in Table D-1 for each RSN sample record, with the exceptions noted below.

As shown in Table D-5, the RSN sample sites in the State of Maine fell in PSU's of two sizes--100 acres and 400 acres. Actually, Maine had a few 200-acre PSU's, but none of these were in the RSN sample. It appears, however, from the RSN sampling documents preserved by the EPA that each 400-acre PSU was treated as four 100-acre PSU's when the RSN sample was selected. The effect of this treatment of 200-acre and 400-acre PSU's in Maine can be seen by considering the factored form of the RSN sampling weight given by (19). It appears from the number of CNI sampling points assigned to the 200-acre and 400-acre PSU's that they were sampled at the same rate as all other PSU's, except for the 640-acre PSU's. Thus, the unconditional CNI factor in the sampling weight (19) is one. The conditional RSN factor is the same, on the average, as that for 100-acre PSU's, since the total points, v, for the 100-acre portion of the 400-acre PSU is same, on the average, as that for 100-acre PSU's. Thus, the mean sampling weight was computed for all sites in Maine as shown for 100-acre PSU's in Table D-1.

The State accumulations of cropland and noncropland ratios, $N_1(i)$ and $N_2(i)$, shown in Table D-6 were checked for logical consistency. This check was felt to be necessary since these values were based upon hard-copy computer output from the RSN sample selection.[3/] This hard-

---

3/ Only a hand written copy could be found for Maine.

Table D-6: State Accumulations* of Cropland Ratios, $N_1(i)$, and Noncropland Ratios, $N_2(i)$, Together with Computed Sample Sizes and Total Land Area**

State Code	State Name	$N_1(i)$	$\hat{n}_1(i)$	$N_2(i)$	$\hat{n}_2(i)$	(1000's of acres) Total Land Area
01	Alabama	467.42123	92	3619.41845	72	32,597
04	Arizona	179.53666	36	9063.14655	176	72,680
05	Arkansas	3656.18536	216[1/]	10472.11792	64	33,468
06	California	1284.14536	268	10698.46741	224	100,076
08	Colorado	1256.23608	240	7390.47656	140	66,486
09	Connecticut	58.31892	8	601.65381	8	3,127
10	Delaware	120.95724	12	154.04031	4	1,266
12	Florida	361.80749	72	3980.34901	80	34,721
13	Georgia	587.92669	120	4066.30930	80	37,263
16	Idaho	655.14071	132	5990.26966	120	52,933
17	Illinois	3050.59595	568	1747.62573	32	35,766
18	Indiana	1616.49756	312	1398.49561	28	23,132
19	Iowa	3160.09155	608	1515.03809	28	35,839
20	Kansas	3426.67927	684	3131.41689	64	52,425
21	Kentucky	718.47949	124	2969.72778	52	25,511
22	Louisiana	546.63631	108	3060.22173	60	28,596
23	Maine	228.28275	32	3735.32919	48	19,848
24	Maryland	423.85901	52	868.16471	12	6,319
25	Massachusetts	56.72992	8	949.29687	12	5,033
26	Michigan	1158.50806	220	3655.05437	68	36,515
27	Minnesota	2557.73877	488	4150.38281	80	51,201
28	Mississippi	627.86577	124	3151.17520	64	30,250
29	Missouri	1702.23242	328	4035.95068	76	44,235

Table D-6: State Accumulations* of Cropland Ratios, $N_1(i)$, and Noncropland Ratios, $N_2(i)$, Together with Computed Sample Sizes and Total Land Area**

State Code	State Name	$N_1(i)$	$\hat{n}_1(i)$	$N_2(i)$	$\hat{n}_2(i)$	(1000's of acres) Total Land Area
30	Montana	1819.92407	340	10579.96484	200	93,098
31	Nebraska	38.16203	40[1/]	1145.83691	120[1/]	49,021
32	Nevada	66.32464	12	8475.24410	176	70,264
33	New Hampshire	57.27583	8	1047.69238	12	5,769
34	New Jersey	157.60745	20	811.71170	12	4,810
35	New Mexico	204.96760	40	9663.70741	192	77,688
36	New York	1212.80716	152	4933.72984	60	30,670
37	North Carolina	694.07227	124	3684.40039	68	31,331
38	North Dakota	3396.55322	636	2544.91626	48	44,442
39	Ohio	1390.60275	276	1889.30123	36	26,206
40	Oklahoma	1310.92419	260	4200.03636	84	43,819
41	Oregon	774.27002	152	7003.43794	140	61,587
42	Pennsylvania	806.77905	152[1/]	3020.83643	56	28,804
44	Rhode Island	7.95469	4	128.59467	4	676
45	South Carolina	378.91528	68	2251.41602	40	19,338
46	South Dakota	2268.04346	420[1/]	4287.50000	80	48,612
47	Tennessee	617.46338	112	3029.04150	56	26,444
48	Texas	3732.70468	744	17409.54254	344	168,001
49	Utah	250.68592	48	6318.30648	128	52,722
50	Vermont	142.21658	20	1017.74146	12	5,937
51	Virginia	742.42676	84	4225.02734	56	25,458
53	Washington	879.95784	180	4461.17404	88[1/]	42,616
54	West Virginia	200.94192	24	2878.29848	36	15,402
55	Wisconsin	1425.55981	272	3141.84985	60	35,013
56	Wyoming	365.49951	68	7940.41016	148	62,306

1/ These values differ from the actual sample sizes in Table 1.3.

* Source: Hard copy computer records of the RSN sampling maintained by the EPA Field Studies Branch, Washington, D.C.

** Source: Basic Statistics--National Inventory of Soil and Water Conservation Needs, 1967.

copy record was believed to be the computer record of the final sample selection for each State, but were not verified. The check made was to compute, as described in Section 1.2.2, the cropland and noncropland sample sizes, $\hat{n}_1(i)$ and $\hat{n}_2(i)$, from the accumulations, $N_1(i)$ and $N_2(i)$, and the total land area of each State as shown in Table D-6. The computed sample sizes, $\hat{n}_1(i)$ and $\hat{n}_2(i)$, differ from the actual sample sizes shown in Table 1.3 by no more than four for all States except Arkansas and Nebraska. Small differences in the computed and actual sample sizes can be explained by the fact the total land area for each State that was used to compute the RSN sample sizes did not agree exactly with the figures shown in Table D-6.[4/] The relatively large discrepancies for Arkansas and Nebraska were interpreted as meaning that the accumulations $N_1(i)$ and $N_2(i)$ shown in Table D-6 for these States are incorrect. Thus, the ratios $N_1(i)/n_1(i)$ and $N_2(i)/n_2(i)$ which were used to compute the sampling weights for these two States, from the formulas shown in Table D-1, were the averages of $N_1(i)/n_1(i)$ and $N_2(i)/n_2(i)$ for all other States, except Rhode Island. The Rhode Island data was also excluded from this average because the very small size of Rhode Island resulted in its cropland sample size being rounded up to 4 even though its computed value was approximately one, which deflated the value of $N_1(i)/n_1(i)$ for Rhode Island.

4/ This is evident from hand computations of the RSN sample sizes preserved by the EPA for some States. The source of the land areas actually used is not known.

APPENDIX E

Construction of an Analysis Data File

Construction of an Analysis Data File

The EPA computer records for the Rural Soils Network (RSN) are structured for simple entry of the data from laboratory analyses. For example, a laboratory test that results in less than detectable levels of a category of compounds produces only a single entry into the data file. In order to simultaneously analyze the data for more than one compound it is useful to restructure the data file so that it contains a distinct variable representing the amount detected for each of the compounds to be analyzed. Thus, a SAS[1/] data set with this structure was created for analysis purposes.[2/] The contents of this data set are shown in Exhibit E-1.

Each detection of a pesticide residue for a sample specimen resulted in an entry into the EPA computer record for each of four variables--a Residue Classification Code (RCC), an Individual Residue Code (IRC), an amount, and a unit. It was found that all amounts were in units of parts per million; thus the unit variable was not included in the SAS file constructed. Only specific residues were tested for on a regular basis. These compounds are listed in Table E-1. Other residues may have been tested ocassionally, but such data cannot be used for inferential purposes. Only data for the pesticides shown in Table E-1

1/ Statistical Analysis System (SAS) User's Guide, SAS Institute, Inc., Raleigh, North Carolina, 1979.

2/ For those readers interested in using this data set, it is stored on a user disk at the EPA North Carolina Computing Center. The fully qualified data set name is

CN.EPAROY.SADD.PEST.SASFILE,

and the data is located in the data set member called TOTAL.

Exhibit E-1. Contents of the SAS data set created*

S T A T I S T I C A L A N A L Y S I S S Y S T E M 18:15 FRIDAY, MARCH 6, 1981 1

CONTENTS OF SAS DATA SET IN.INITIAL

TRACKS USED=297 SUBEXTENTS=1 OBSERVATIONS=12372 CREATED BY JOB FPAROY27 AT 18:15 FRIDAY, MARCH 6, 1981

BY SAS RELEASE 79.08 DSNAME=CA.FPAROY.SADD.PEST.SASFILE BLKSIZE=13030 LRECL=307 OBSERVATIONS PER TRACK=42 GENERATED BY DATA

ALPHABETIC LIST OF VARIABLES

#	VARIABLE	TYPE	LENGTH	POSITION	FORMAT	INFORMAT	LABEL
1	ACCNUM	NUM	4	4			ACCESSION NUMBER
18	ANDATE	NUM	4	52			ANALYSIS DATE
15	CLAY	NUM	3	44			
70	CONAME	CHAR	20	272			COUNTY NAME
3	COUNTY	NUM	3	10			
11	CROPNUM	NUM	2	34			CROP NUMBER
17	CROPREG	NUM	2	50			CROPPING REGION
14	CROPYR	NUM	2	56			CROP YEAR
20	FY	NUM	2	58			FISCAL YEAR
7	LAB	NUM	2	21			
6	LANDUSE	NUM	2	19			LAND USE
16	ORGMAT	NUM	3	47			
22	PEST1002	NUM	4	64			ALDRIN
67	PEST1013	NUM	4	244			ARSENIC
68	PEST1016	NUM	4	248			ATRAZINE
44	PEST1080	NUM	4	152			BENZENE HEPTACHLORIDE
55	PEST1091	NUM	4	196			BUTAN
57	PEST1149	NUM	4	204			CARBOPHENOTHION
23	PEST1160	NUM	4	68			CHLORDANE
56	PEST1161	NUM	4	200			GAMMA CHLORDANE
66	PEST1235	NUM	4	240			2,4-D
24	PEST1237	NUM	4	72			DCPA
25	PEST1240	NUM	4	76			O,P-'DDE
26	PEST1241	NUM	4	80			P,P-'DDE
27	PEST1243	NUM	4	84			P,P-'DDT
28	PEST1244	NUM	4	88			O,P-'DDT
58	PEST1246	NUM	4	208			DEF
59	PEST1248	NUM	4	212			DIAZINON
31	PEST1258	NUM	4	100			DICOFOL
32	PEST1260	NUM	4	104			DIELDRIN
34	PEST1336	NUM	4	112			ENDOSULFAN I
35	PEST1337	NUM	4	116			ENDOSULFAN II
36	PEST1338	NUM	4	120			ENDOSULFAN SULFATE
37	PEST1341	NUM	4	124			ENDRIN
38	PEST1342	NUM	4	128			ENDRIN ALDEHYDE
39	PEST1343	NUM	4	132			ENDRIN KETONE
60	PEST1348	NUM	4	216			ETHION
61	PEST1380	NUM	4	220			FOLEX
40	PEST1420	NUM	4	136			HEPTACHLOR
41	PEST1421	NUM	4	140			HEPTACHLOR EPOXIDE
42	PEST1448	NUM	4	144			ISODRIN
43	PEST1457	NUM	4	148			LINDANE
21	PEST1499	NUM	4	60			ALACHLOR
62	PEST1518	NUM	4	224			MALATHION
45	PEST1526	NUM	4	156			METHOXYCHLOR
63	PEST1531	NUM	4	228			METHYL PARATHION
51	PEST1534	NUM	4	180			MIREX
52	PEST1620	NUM	4	184			OVEX

*Source: Computer files supplied by the EPA Field Studies Branch, Washington, D.C.

Exhibit 1 (continued)

33	PEST638	NUM	4	108	PHOTODIELDRIN
64	PEST643	NUM	4	232	ETHYL PARATHION
46	PEST646	NUM	4	160	PCNB
65	PEST650	NUM	4	236	PHORATE
54	PEST670	NUM	4	192	PROLAN
53	PEST672	NUM	4	188	PCB
47	PEST687	NUM	4	164	PROPACHLOR
48	PEST688	NUM	4	168	RONNEL
29	PEST786	NUM	4	92	P,P-'DDE
30	PEST787	NUM	4	96	O,P-'DDE
49	PEST795	NUM	4	172	TOXAPHENE
50	PEST799	NUM	4	176	TRIFLURALIN
12	PH	NUM	3	35	
9	RAIN	NUM	3	27	
72	ROUND	NUM	2	294	SAMPLE ROUND NO.(1 OR 2)
8	SAMPDATE	NUM	4	23	SAMPLING DATE
13	SAND	NUM	3	38	
14	SILT	NUM	3	41	
4	SITE	NUM	3	13	
5	SMC	NUM	3	16	SAMPLE MATERIAL CODE
2	STATE	NUM	2	8	
69	STNAME	CHAR	20	252	STATE NAME
73	STRATUM	NUM	3	296	LARGE SCALE GEO. STRATUM ID
10	TEMP	NUM	3	30	
74	WT	NUM	8	299	SAMPLING WEIGHT
71	YEAR	NUM	2	292	SAMPLE YEAR (1,2,3 OR 4) IN ROUND

```
+------------------------------ SOURCE STATEMENTS ------------------------------------+
\DATA IN.TOTAL; SET ONE;                                                       8.   \
+-------------------------------------------------------------------------------------+
```

Table E-1: Pesticide Residues Tested on a Regular Basis*

Residue Classification Code (RCC)	Individual Residue Code (IRC)	Compound
2	499	Alachlor
2	002	Aldrin
2	160	Chlordane
2	237	DCPA
2	240	o,p - 'DDE
2	241	p,p - 'DDE
2	243	p,p - 'DDT
2	244	o,p - 'DDT
2	786	p,p - 'TDE
2	787	o,p - 'TDE
2	258	Dicofol
2	260	Dieldrin
2	638	Photodieldrin
2	336	Endosulfan I
2	337	Endosulfan II
2	338	Endosulfan Sulfate
2	341	Endrin
2	342	Endrin Aldehyde
2	343	Endrin Ketone
2	420	Heptachlor
2	421	Heptachlor Epoxide
2	448	Isodrin
2	497	Lindane
2	080	Benzene Heptachloride
2	526	Methoxychlor
2	646	PCNB
2	687	Propachlor
2	688	Ronnel
2	795	Toxaphene
2	799	Trifluralin
2	534	Mirex
2	620	Ovex
2	672	PCB
2[1/]	670[2/]	Prolan
2[1/]	091[2/]	Bulan
2	161[3/]	Gamma Chlordane

(cont.)

1/ Shown on the computer records to have RCC = 7, but corrected to 2 by personal communication with EPA Field Studies Branch, Washington, D.C.

2/ Tested for fiscal year 1972 specimens, and thereafter.

3/ Tested for fiscal year 1974 specimens, and thereafter.

Table E-1: Pesticide Residues Tested on a Regular Basis* (continued)

Residue Classification Code (RCC)	Individual Residue Code (IRC)	Compound
3	149	Carbophenothion
3	246	DEF
3	248	Diazinon
3	348	Ethion
3	380	Folex
3	518	Malathion
3	531	Methyl Parathion
3	643	Ethyl Parathion
3	650	Phorate
4	235	2, 4-D
5	013 4/	Arsenic
6	016 4/	Atrazine

4/ Tested in fiscal years 1969, 1972, and 1973 for specimens from "cornbelt States" only, i.e., South Dakota, Nebraska, Kansas, Missouri, Iowa, Minnesota, Wisconsin, Illinois, Indiana, Ohio, and Michigan.

* Source: Personal communication from the EPA Field Studies Branch, Washington, D.C.

Table E-2: Residue Classification Codes*

Residue Classification Code (RCC)	IRC for "none found"	Compound Category
2	905	Chlorinated hydrocarbons
3	910	Organophosphorous insecticides
4	911	Phenoxy acid derivative herbicides
5	901	Arsenic compounds
6	914	Triazines

* Source: Personal communication with EPA Field Studies Branch, Washington, D.C.

were included in the SAS data file. Table E-1 also shows the RCC for each of the compounds tested regularly, and Table E-2 gives the discription of each of these RCC categories. The RCC categories are crucial to proper analysis of the data because all compounds with a common RCC are tested simultaneously. If the test is performed for compounds with RCC = 2, for example, there are two possible types of entry into the computer record. Either each positive detection is entered, or an IRC code is entered to indicate no positive detections, as shown in Table E-2. Each record in the EPA computer file corresponds to a specific sample specimen and contains 40 repitions of fields for IRC, RCC, amount, and unit. All of these fields were replaced in the SAS data set by 48 pesticide amount variables, one for each of the 48 compounds listed in Table E-1. A zero was entered for each compound for which less than detectable levels were found.[2/] A decimal point, the SAS

[2/] Indicated by one or more detectable amounts for the same RCC or a 9XX code as shown in Table E-2.

missing value symbol, was entered for the amount of compound detected whenever the test for that compound was not performed.[3/]

There are several variables, included for analysis purposes, on the SAS data file that were not on the original EPA data file. Among these new variables are STNAME and CONAME, the State and county names. A variable called ROUND was constructed which has a value of one for records from the first round of data collection, and a value of two for the second round when the sites were revisited. Also, the sampling year within round is given by YEAR, e.g., YEAR = 1 for first-year sample sites. If the information were available, it would also be useful to have an indicator variable to identify when site substitutions were made. This is especially important when substitutions were made in the second round; the second round data for such a site should not be directly compared to the first round data for that site.[4/] This is an important consideration when estimating differences in residue levels from the first round to the second round.

Two other variables added to the data file for analysis purposes are STRATUM and WT. The variable STRATUM is used to identify large-scale geographic strata within States as described in section 1.7.7. The STRATUM codes and their meanings are given in Table E-3. The variable WT is the approximate sampling weight, which was constructed as

---

[3/] Indicated by no detectible amounts for the same RCC and no corresponding 9XX code as shown in Table E-2.

[4/] EPA Field Studies Branch, Washington, D.C., assured RTI that such substitutions in the second round amount to no more than 5 percent of all second round sites.

Table E-3: STRATUM Codes and Their Meaning*

STRATUM Code	Meaning
40	Irrigated stratum
100 or 160[1/]	Remainder stratum[2/]
400 or 640[1/]	Sandhills, desert, or other relatively homogeneous stratum

1/ Code used depends on PSU sizes used in the State.

2/ All sites in many states fall into the remainder stratum.

* Source: Constructed by RTI from: a) Data files supplied by the EPA Field Studies Branch, Washington, D.C. b) Data and personal communications from J. Jeffery Goebel, Statistical Laboratory, Iowa State University.

shown in Appendix D. This variable is, of course, essential for a weighted analysis of the data that incorporates the sampling design implications.

Some quality assurance checks of the EPA computer files for the RSN were made prior to creation of the SAS data set for analysis. Twenty-three inconsistencies were discovered. These inconsistencies are summarized in Table E-4, and their resolution is discussed below. Most of these inconsistencies were resolved by consulting microfilm copies of the Analysis Worksheets, Form 6-7, and the Sample Data Sheets, Form 6-4.[5/]

5/ Maintained by the EPA Field Studies Branch, Washington, D.C.

Table E-4: Data Inconsistancies in the Rural Soils Network Files*

Case Number	State Name (State Number)	Site Number	Fiscal Year	Sample Material Code (SMC)	Accession Number	Individual Residue Codes (IRC)	Residue Class Codes (RCC)
1	California (06)	39	69	1	3196 3407	13,244,243,241,260,786 911	2,5 4
2	Idaho (16)	67	69	1	1470 11470	905 13	2 5
3	Idaho (16)	75	69	1	1471 11471	160,260 13	2 5
4	Missouri (29)	21	69	1	1226 11226	13 905	5 2
5	North Carolina (37)	31	69	1	4063 4061	13,241,786,243,240,787 911	2,5 4
6	Ohio (39)	22	69	1	3566 3568	241,260,2,243 13	2 5
7	Virginia (51)	37	69	1	3468 4049	13,905 911	2,5 4
8	Virginia (51)	46	69	1	795 3476	13,911 905	4,5 2
9	Illinois (17)	138	69	1	3017 3017[1/]	13,160,914 911	2,5,6 4
10	New York (36)	78	70	63	10049 100049	905 910	2 3
11	Alabama (01)	91	72	1	204007 204117	13,241,243,910 914	2,3,5 6
12	Mississippi (28)	113	72	1	204298 204298	241,244,243 13	2 5
13	Iowa (19)	559	73	1	312655 372655	799 910	2 3
14	Oregon (41)	103	73	1	310110 316110	910 905	3 2

(continued)

Table E-4: Data Inconsistancies in the Rural Soils Network Files* (continued)

Case Number	State Name (State Number)	Site Number	Fiscal Year	Sample Material Code (SMC)	Accession Number	Individual Residue Codes (IRC)	Residue Class Codes (RCC)
15	Pennsylvania (42)	164	73	1	314025 340250	16,241,243 910	2,6 3
16	Nebraska (31)	151	74	1	426298 427298	260 910	2 3
17	Illinois (17)	99	69	1	3028 3017[1/]	13,905 911	2,5 4
18	Louisiana (22)	26	69	1	3621 4365	13,2,260 910	2,5 3
19	Mississippi (28)	49	70	138	8652 8675	795,244,243,241,786,910 241,244,243,795,246	2,3 2,3
20	Alabama (01)	105	72	1	204111 204014	13,241,910 13,260,910	2,3,5 2,3,5
21	New York (36)	194	73	1	314144 314085	240,244,786,241,243,914 910	2,6 3
22	West Virginia (54)	54[2/]	73	1	314097	905,910	2,3
23	Mississippi (28)	49	69	1	781	13,243,244,786,241,341, 795,246,799,240	2,5[3/]

[1/] Case 17 becomes case 9 after the site number for case 17 is corrected to 138.

[2/] Noncropland site number; land use changed from cropland (1) to noncropland (2).

[3/] RCC8 changed to 3 as IRC8 = 246. See Table E-1.

* Source: Computer files supplied by EPA Field Studies Branch, Washington, D.C.

One final correction to the data file was to correct the cropping region code for several counties. Valid codes for the cropping regions are the integers from one through 8. Several records in the computer file showed cropping region codes of 0 and 9. These records were corrected as shown in table E-5.

Table E-5. Resolution of Invalid Cropping Region Codes*

State Name (State Number)	County Name (County Number)	Cropping Region	
		Original	Corrected
Iowa (19)	Scott (163)	0	1
Kentucky (21)	Scott (209)	0	6
Minnesota (27)	Scott (139)	0	5
Mississippi (28)	Alcorn (3)	0	3
Missouri (29)	Scott (201)	0	4
Nebraska (31)	Hayes (85)	0	2
Nebraska (31)	Scotts Bluff (157)	0	5
Nebraska (31)	Thayer (169)	0	1
Oklahoma (33)	Cotton (33)	0	2
South Carolina (35)	Dorchester (35)	0	4
Tennessee (47)	Haywood (75)	0	3
Tennessee (47)	Scott (151)	0	6
California (6)	Alpine (3)	9	Missing
Georgia (13)	Invalid (4)	9	Missing
Maryland (24)	Invalid (18)	9	Missing
New York (36)	Bronx (5)	9	Missing
New York (36)	Invalid (32)	9	Missing
New York (36)	New York (61)	9	Missing
North Carolina (37)	Dare (55)	9	Missing
Virginia (51)	Invalid (39)	9	Missing
Virginia (51)	Invalid (74)	9	Missing
Virginia (51)	Norfolk (129)	9	Missing
Virginia (51)	Princess Anne (151)	9	Missing
West Virginia (54)	McDowell (47)	9	Missing

*Personal communication with EPA Field Studies Branch, Washington, D.C.